

STIC-Biotech/ChemLib

123121

**From:** Chan, Christina  
**Sent:** Thursday, May 27, 2004 12:18 PM  
**To:** Schnizer, Holly; STIC-Biotech/ChemLib  
**Subject:** RE: Request for RUSH sequence search in Appl. no. 09/902772

Please rush. Thanks Chris

Chris Chan

TC 1600 New Hire Training Coordinator and SPE 1644 & 1642  
(571)-272-0841  
Remsen, 3E89

CRFE

-----Original Message-----

**From:** Schnizer, Holly  
**Sent:** Thursday, May 27, 2004 12:14 PM  
**To:** Chan, Christina  
**Subject:** Request for RUSH sequence search in Appl. no. 09/902772

I would like to request a RUSH sequence search for the sequences below in the above appl. which is an after-final.

Please search all databases for

SEQ ID NO:1  
reverse translation of SEQ ID NO:2.

Thank you.

Holly Schnizer  
AU 1653  
Office: REM 3C79  
Mailbox: REM 3C70  
(571)272-0958

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: \_\_\_\_\_  
Date Completed: 5/28  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:

NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)

STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: May 27, 2004, 16:28:06 ; Search time 5914 Seconds  
(without alignments)  
10787.301 Million cell updates/sec  
Title: US-09-902-772-1  
Perfect score: 1447  
Sequence: 1 gaattccgcgaacgaataat.....gaaagaagcgccaagaaaa 1447

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 3470272 seqs, 21671516995 residues  
Total number of hits satisfying chosen parameters: 6940544  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:  
1: gb\_ba:\*  
2: gb\_hgt:\*  
3: gb\_in:\*  
4: gb\_on:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_on:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vi:\*  
30: em\_hgt\_hum:\*  
31: em\_hgt\_inv:\*  
32: em\_hgt\_other:\*  
33: em\_hgt\_mus:\*  
34: em\_hgt\_pln:\*  
35: em\_hgt\_rtd:\*  
36: em\_hgt\_mam:\*  
37: em\_hgt\_vrt:\*  
38: em\_sy:\*  
39: em\_hgtgo\_hum:\*  
40: em\_hgtgo\_mus:\*  
41: em\_hgtgo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1447	100.0	1447	6	E31253	E31253 Protein hav
2	1409	97.4	1413	5	AY065661	AY065661 Gallus ga
3	1356	93.7	1528	6	E31254	E31254 Protein hav
4	1344	92.9	1516	5	GGERG	X77159 G.gallus ER
5	994.4	68.7	1509	6	AX657228	AX657228 Sequence
6	953.4	65.9	2016	9	BC040168	BC040168 Homo sapi
7	936	64.7	1808	10	AB031088	AB031088 Rattus no
8	921	63.6	2166	9	HUMERG2	M17254 Human erg2
9	889.6	61.5	2012	5	XLAJA4126	AJ224136 Xenopus 1
10	881.6	60.9	2131	10	AB073078	AB073078 Mus muscu
11	867.2	59.9	2209	10	AB073080	AB073080 Mus muscu
12	833.6	57.6	1291	6	AX657230	AX657230 Sequence
13	828	57.2	2133	10	AB073079	AB073079 Mus muscu
14	801.8	55.4	1798	5	XLAJA4125	AJ224135 Xenopus 1
15	760.2	52.5	3126	9	HUMERG11	M21535 Human erg p
16	742.6	51.3	1372	6	AX657234	AX657234 Sequence
17	698.8	48.3	1219	6	AX657232	AX657232 Sequence
18	571.2	39.5	1436	5	DRE249590	AJ249590 Danio rer
19	570.2	39.4	1359	6	AX695580	AX695580 Sequence
20	570.2	39.4	1729	6	AX695579	AX695579 Sequence
21	570.2	39.4	1729	10	MMFLI1	X59421 Mouse Fli-1
22	563.8	39.0	2916	6	A36461	A36461 Sequence 2
23	563.8	39.0	2938	6	AR080101	AR080101 Sequence
24	562.2	38.9	2388	9	HSUMFLI	X67001 H.sapiens H
25	562.2	38.9	1359	6	AX695583	AX695583 Sequence
26	562.2	38.9	2394	9	AY029368	AY029368 Homo sapi
27	562.2	38.9	2394	6	BC010115	BC010115 Homo sapi
28	562.2	38.9	2957	6	AX587825	AX587825 Sequence
29	562.2	38.9	2957	6	AX695582	AX695582 Sequence
30	562.2	38.9	2957	9	HUMERGBFLI	M98833 Homo sapien
31	562.2	38.9	2959	9	BC001670	BC001670 Homo sapi
32	561.4	38.8	6849	12	AY029367	AY029367 Synthetic
33	560.6	38.7	1673	9	S45205	S45205 Fli-1-Frien
34	549.8	38.0	2919	5	BC056588	BC056588 Danio rer
35	549.4	38.0	4403	5	XLFLIG	X66979 X.laevis mr
36	544.6	37.6	2954	6	BD094081	BD094081 Shear str
37	518.6	35.8	1245	5	AF177538	AF177538 Danio rer
38	514	35.5	3490	5	CCRNAPLI	Y14773 Coturnix co
39	502.2	34.7	3545	5	CCFLIONCO	Y14774 Coturnix co
40	493.2	34.1	762	9	S72621	S72621 EWS...erg {
41	478.2	33.0	1932	9	HUMFLI1A	M93255 Human Fli-1
42	477	33.0	816	10	S66169	S66169 Erg-3-immun
43	388.6	26.9	127477	9	BS000128	BS000128 Pan trogl
44	388.6	26.9	148367	9	BS000159	BS000159 Pan trogl
45	387	26.7	14518	9	AP001423	AP001423 Homo sapi

ALIGNMENTS

RESULT 1  
E31253  
LOCUS  
DEFINITION Protein having cell calcifying inhibitory activity and gene  
encoding the same.  
ACCESSION E31253  
VERSION E31253.1 GI:13025685  
KEYWORDS JP 199075871-A/1.  
SOURCE unidentified  
ORGANISM unclassified  
REFERENCE 1 (bases 1 to 1447)  
AUTHORS Hiroyasu, I., Yoshinobu, H., Maritjo, P., Joel, R. and Helena, E.  
TITLE Protein having cell calcifying inhibitory activity and gene  
encoding the same

ALIGNMENTS  
E31253 1447 bp DNA linear PAT 18-JUN-2001  
LOCUS Protein having cell calcifying inhibitory activity and gene  
encoding the same.  
ACCESSION E31253  
VERSION E31253.1 GI:13025685  
KEYWORDS JP 199075871-A/1.  
SOURCE unidentified  
ORGANISM unclassified  
REFERENCE 1 (bases 1 to 1447)  
AUTHORS Hiroyasu, I., Yoshinobu, H., Maritjo, P., Joel, R. and Helena, E.  
TITLE Protein having cell calcifying inhibitory activity and gene  
encoding the same

Pred. No. is the number of results predicted by chance to have a

JOURNAL	Patent: JP 199075871-A 1 23-MAR-1999;
COMMENT	CHUGAI PHARMACEUT CO LTD, HE TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA
OS	Unidentified
PN	JP 199075871-A/1
PD	23-MAR-1999
PF	29-MAY-1998 JP 1998166076
PR	18-JUN-1997 US 08/878177, 20-JUN-1997 US 60/050297, PI HIROYASU IWAMOTO, YOSHINOBU HIGUCHI, MARIJIO FASHIFIKI, PI JOEL ROZENBUROOM,
PI	HELENA E
PC	C12N15/09, A61K48/00, C07K14/465, C07K16/18, C12Q1/68//A61K38/00,
PC	C12P21/02,
PC	C12P21/08, C12N15/00, A61K37/02
CC	Strandedness: Single;
CC	Topology: Linear;
PH	Key
FT	source
FT	1. .1447
FEATURES	Location/Qualifiers
source	1. .1447
	/organism="unidentified"
	/mol_type="genomic DNA"
	/db_xref="taxon:32644"
ORIGIN	
Query Match	100.0%; Score 1447; DB 6; Length 1447;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 1447; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Qy	1 GAATTCGCGAAGCAATTAATTATTATTAGCAATATTATAGCATCAATAATCTTGATCACA 60
Db	1 GAATTCGCGAAGCAATTAATTATTATTAGCAATATTATAGCATCAATAATCTTGATCACA 60
Qy	61 TTATGGCAAGCACTATTAAAGAACATATACAGTGTGAGTGAGACCAAGTCTTGTTG 120
Db	61 TTATGGCAAGCACTATTAAAGAACATATACAGTGTGAGTGAGACCAAGTCTTGTTG 120
Qy	121 AGTGTGCTTACGGATTGCGCCACCTTGCCAAAGACAGAAATGACAGAGCTCTCTTCCAGTG 180
Db	121 AGTGTGCTTACGGATTGCGCCACCTTGCCAAAGACAGAAATGACAGAGCTCTCTTCCAGTG 180
Qy	181 AATATGGCAAAATCATCAAGATGAGCCGCGCTTCCAGAGGAGTCTGGTTATCACAGC 240
Db	181 AATATGGCAAAATCATCAAGATGAGCCGCGCTTCCAGAGGAGTCTGGTTATCACAGC 240
Qy	241 CCCCGCCAGAGTTACCATTAAGATGGAGTGTACCCACACACAGGTTAATGGTCAAGA 300
Db	241 CCCCGCCAGAGTTACCATTAAGATGGAGTGTACCCACACACAGGTTAATGGTCAAGA 300
Qy	301 ATTCACTGATGACTGCAGCGTGGCAAAGAGAGGAAATGGTTAGCAGTTCAGACAATG 360
Db	301 ATTCACTGATGACTGCAGCGTGGCAAAGAGAGGAAATGGTTAGCAGTTCAGACAATG 360
Qy	361 TTGGGATGAACCTATGGAGCTACATGGAGAGAGGACATATTCGCGCTCCAAATATGACAA 420
Db	361 TTGGGATGAACCTATGGAGCTACATGGAGAGAGGACATATTCGCGCTCCAAATATGACAA 420
Qy	421 CCAATCAAGCAAGAGTTATTGTGCCAGAGATCCTACGTTATGGAGCAGACCAATGTAC 480
Db	421 CCAATCAAGCAAGAGTTATTGTGCCAGAGATCCTACGTTATGGAGCAGACCAATGTAC 480
Qy	481 GGCAGTGGCTGGAGTGGGAGTGAAGAGTATGCTCTTCCAGAGTGGACATCTTGTTG 540
Db	481 GGCAGTGGCTGGAGTGGGAGTGAAGAGTATGCTCTTCCAGAGTGGACATCTTGTTG 540
Qy	541 TCCAGAACATTGATGGGAAGAGTTGTGTAAATATGACCAAGATGACTTCCAGAGCTCA 600
Db	541 TCCAGAACATTGATGGGAAGAGTTGTGTAAATATGACCAAGATGACTTCCAGAGCTCA 600
Qy	601 CGCGAGCTATAACGAGATATCTCTCTGTGCACCTACTACTCTCAGAGAGAGAGAG 660
Db	601 CGCGAGCTATAACGAGATATCTCTCTGTGCACCTACTACTCTCAGAGAGAGAGAG 660

Yeh, H., Abrams, W.R., Rosenbloom, J. and Pacifici, M.  
Transcription factor ERG variants and functional diversification of  
Chondrocytes during limb long bone development  
J. Cell Biol. 150 (1), 27-40 (2000)  
20351415  
PUBMED  
REFERENCE  
2 10893254  
AUTHORS  
Iwamoto, M., Higuchi, Y., Enomoto-Iwamoto, M., Kurisu, K., Koyama, E.,  
Yeh, H., Rosenbloom, J. and Pacifici, M.  
The role of ERG (ets related gene) in cartilage development  
Osteoarthritis. Cartil. 9 Suppl A, S41-S47 (2001)  
21535378  
PUBMED  
REFERENCE  
3 11680687  
AUTHORS  
Iwamoto, M., Higuchi, Y., Yeh, H. and Pacifici, M.  
Direct Submission  
Submitted (05-DEC-2001) Oral Anatomy and Developmental Biology,  
Osaka University Faculty of Dentistry, 1-8 Yamadaoka, Suita, Osaka  
565-0871, Japan  
FEATURES  
source  
1. 1413  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9031"  
37. 11392  
/note="alternatively spliced; transcription factor"  
/codon\_start=1  
/product="erg isoform C-1-1"  
/protein\_id="AA140899.1"  
/db\_xref="GI:17887441"  
/translation="MSTIKELSVSEDSLEPCAYGSPHLAKTENTASSSEYQQT  
SKMSPRVQDNLSPQPARVIRKMEFNQVNSRNSPDDCSVAKGKMVSSSDNVGM  
NYGSEMEKHIPPNMTNERRVVPADTLSTWDRQLEWAVKEYGLPVDVILLF  
QNTDGKELKMTKDDFORLFTSYNADILLSHLYLRERGAFFPNTSVYPEATORIT  
TRPLPYEQARRSAWTSHTSPTOSKATPSSSTVPKTEDORPOLDPQILGPTSSRLA  
NPGSGOIQLOFLELLLSNNSNCITWEGTNGEFKMTDPEVARWGRKSKPNNVY  
DKLSRALRYVDKNNIKVGRYKVFDFHGIQAOLQHPHPRESSVVKYPSDLPYMS  
YHAPQKGNFVAPHPALPVTSSSFPAAPYWNPSPTGGIYFNTRLPAAHMFSLHTY  
Y"  
ORIGIN  
Query Match 97.4%; Score 1409; DB 5; Length 1413;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1409; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
27 TAGCAATTATPAGCGATCAATATCTTGTATGACATTAATGCGAGCACTATAAGGAAGCA 86  
Db 1 TAGCAATTATPAGCGATCAATATCTTGTATGACATTAATGCGAGCACTATAAGGAAGCA 60  
87 TTATCAGTGTGAGTGAAGCAGTCTCTTGTTCAGTGTGCTTACGATCGCCCACTT 146  
Db 61 TTATCAGTGTGAGTGAAGCAGTCTCTTGTTCAGTGTGCTTACGATCGCCCACTT 120  
147 GCAAGACAGAAATGACAGCCTCTCTTCCAGTGAATATGCGGCAACATCAAGATGAGC 206  
Db 121 GCAAGACAGAAATGACAGCCTCTCTTCCAGTGAATATGCGGCAACATCAAGATGAGC 180  
207 CCGCGCTTCCCGAGCAGGAGTGTATACAGCCCGCGGCGAGTACCATTAAGATG 266  
Db 181 CCGCGCTTCCCGAGCAGGAGTGTATACAGCCCGCGGCGAGTACCATTAAGATG 240  
267 GAGTGTAAACCAACAGGTTAAATGGGTCAAGGAATTCACCTGATGCTGCGAGCGTGGCA 326  
Db 241 GAGTGTAAACCAACAGGTTAAATGGGTCAAGGAATTCACCTGATGCTGCGAGCGTGGCA 300  
327 AAGGAGGGAATGGTTAGCAGTTCAGACAAATGTTGGATGACTATGAAAGCTACATG 386  
Db 301 AAGGAGGGAATGGTTAGCAGTTCAGACAAATGTTGGATGACTATGAAAGCTACATG 360  
387 GAAGAGAGCATATTCGCCCTCCAAATATGACCAATGAAACGAAGAGTTATTGTGCCA 446  
Db 361 GAAGAGAGCATATTCGCCCTCCAAATATGACCAATGAAACGAAGAGTTATTGTGCCA 420

QY 447 GCAGATCCTACGTTTATGAGCAGACACCATGTACGGCAGTGGCTGAGTGGGCACTGAAG 506  
Db 421 GCAGATCCTACGTTTATGAGCAGACACCATGTACGGCAGTGGCTGAGTGGGCACTGAAG 480  
QY 507 GAGTATGCTTCCAGAGCGTGAACATCTTGTGTTCAGAAACATTCATGATGGGAAAGAGTTG 566  
Db 481 GAGTATGCTTCCAGAGCGTGAACATCTTGTGTTCAGAAACATTCATGATGGGAAAGAGTTG 540  
QY 567 TGTAAATGACCAAGATGACTTCCAGAGACTTCAGCGGAGCTACCGCGAGCTATAACGCGAGATATCCTC 626  
Db 541 TGTAAATGACCAAGATGACTTCCAGAGACTTCAGCGGAGCTACCGCGAGCTATAACGCGAGATATCCTC 600  
QY 627 CTGTCCACACCTACACTACCTCAGAGAGAGAGAGCCACTTTTATTTTCCAAATACATCA 686  
Db 601 CTGTCCACACCTACACTACCTCAGAGAGAGAGAGCCACTTTTATTTTCCAAATACATCA 660  
QY 687 GTTTACCCAGAGACCAACCAAGATACACAGAGCCAGATTTTACCTTATGAGCAAGCG 746  
Db 661 GTTTACCCAGAGACCAACCAAGATACACAGAGCCAGATTTTACCTTATGAGCAAGCG 720  
QY 747 AGGAGATCAGCGTGGACGAGTCAACAGCCATCCCCTCAGTCAAAAGCTACCCCAACATCA 806  
Db 721 AGGAGATCAGCGTGGACGAGTCAACAGCCATCCCCTCAGTCAAAAGCTACCCCAACATCA 780  
QY 807 TCTTCAACAGTCCCAAAACAGAGACAGCGTCTCTCAGTTAGATCCTTATCAGATTTCTT 866  
Db 781 TCTTCAACAGTCCCAAAACAGAGACAGCGTCTCTCAGTTAGATCCTTATCAGATTTCTT 840  
QY 867 GGACCGACAGCAGCGCTCTTGCATTCAGGAGTGGGCGAGTACAGTATGGCAGTTTC 926  
Db 841 GGACCGACAGCAGCGCTCTTGCATTCAGGAGTGGGCGAGTACAGTATGGCAGTTTC 900  
QY 927 CTATCGAGCTTCTCTCGGACAGCTCCAACTCAACCTGATCCTTACCTGGAGGGCAAAAT 986  
Db 901 CTATCGAGCTTCTCTCGGACAGCTCCAACTCAACCTGATCCTTACCTGGAGGGCAAAAT 960  
QY 987 GGGGAGTTCAAGATCAGACAGCCTCATGAAGTGGCTCGCGTGGGGAGAGAGGAAAGC 1046  
Db 961 GGGGAGTTCAAGATCAGACAGCCTCATGAAGTGGCTCGCGTGGGGAGAGAGGAAAGC 1020  
QY 1047 AAACCTAACATGAATATGACAACTCAGCGCTGCACTTCGCTACTACTATGACAAAT 1106  
Db 1021 AAACCTAACATGAATATGACAACTCAGCGCTGCACTTCGCTACTACTATGACAAAT 1080  
QY 1107 ATTATGACTAAAGTTTCATGTAAGAGCTATGCTCAAAATTTGATTTCACCGAATCGCT 1166  
Db 1081 ATTATGACTAAAGTTTCATGTAAGAGCTATGCTCAAAATTTGATTTCACCGAATCGCT 1140  
QY 1167 CAGGCGCTCCAGCTCAGCTCCAGAAATCATCCATGTACAAATACCCATCAGACCTCCCC 1226  
Db 1141 CAGGCGCTCCAGCTCAGCTCCAGAAATCATCCATGTACAAATACCCATCAGACCTCCCC 1200  
QY 1227 TACATGAGTTCCTACCATGCACACCCCGAGAGATGAATTTGTAGTCCCATCCCCCT 1286  
Db 1201 TACATGAGTTCCTACCATGCACACCCCGAGAGATGAATTTGTAGTCCCATCCCCCT 1260  
QY 1287 GTTTGCGCGTCAACCTCATCCAGCTTTTGTGTCCTTAATPCCATCTGGAATCAACA 1346  
Db 1261 GTTTGCGCGTCAACCTCATCCAGCTTTTGTGTCCTTAATPCCATCTGGAATCAACA 1320  
QY 1347 ACTGAGGCACTACCCCAATACAGGCTGCGAGTGTCTCATATGCTTCCCATCTTGGC 1406  
Db 1321 ACTGAGGCACTACCCCAATACAGGCTGCGAGTGTCTCATATGCTTCCCATCTTGGC 1380  
QY 1407 ACCTACTACTAAGTGGGAAAGAAAGAAA 1435  
Db 1381 ACCTACTACTAAGTGGGAAAGAAAGAAA 1409

RESULT 3  
E31254  
LOCUS  
DEFINITION Protein having cell calcifying inhibitory activity and gene  
E31254 1528 bp DNA linear PAR 19-JUN-2001



```

encoding the same.
E31254
VERSION E31254.1 GI:13025686
KEYWORDS JP 1999075871-A/2.
SOURCE unidentified
ORGANISM unclassified.
1 (bases 1 to 1528)
REFERENCE Hiroyasu,I., Yoshinobu,H., Marijo,P., Joel,R. and Helena,E.
AUTHORS Protein having cell calcifying inhibitory activity and gene
TITLE encoding the same
JOURNAL Patent: JP 1999075871-A 2 23-MAR-1999;
CHUGAI PHARMACEUT CO LTD, HE TRUSTEES OF THE UNIVERSITY OF
PENNSYLVANIA
OS Unidentified
PN JP 1999075871-A/2
PD 23-MAR-1999
PF 29-MAY-1998 JP 1998166076
PR 18-JUN-1997 US 08/878177, 20-JUN-1997 US 60/050297 PI
HIROYASU IWAMOTO, YOSHINOBU HIGUCHI, MARIJO PASHIFIKI, PI JOEL
ROZENBUROOM
PI HELENA E
PC C12N15/09,A61K48/00,C07K14/465,C07K16/18,C12Q1/68//A61K38/00,
C12P21/02,
PC C12P21/08,C12N15/00,A61K37/02
CC Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1..1528 /organism="Unidentified".
FT Location/Qualifiers
1..1528
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"

Query Match 93.7%; Score 1356; DB 6; Length 1528;
Best Local Similarity 94.7%; Pred. No. 0;
Matches 1447; Conservative 0; Mismatches 0; Indels 81; Gaps 1;

QY 1 GAATTCGCGAAGCAATTAATTATTAGCAATTAATTAGCGATCAATAATCTTGATCA 60
DB 1 GAATTCGCGAAGCAATTAATTATTAGCAATTAATTAGCGATCAATAATCTTGATCA 60
QY 61 TTATGGCAAGCACTATTAAAGGAAGCAATTAATCAAGTGGTGAAGACCACTCTGTTG 120
DB 61 TTATGGCAAGCACTATTAAAGGAAGCAATTAATCAAGTGGTGAAGACCACTCTGTTG 120
QY 121 AGTGTGCTACGGATCGCCCACTTCGAAAGACAGAAATGACAGCCCTCTCTCCAGTG 180
DB 121 AGTGTGCTACGGATCGCCCACTTCGAAAGACAGAAATGACAGCCCTCTCTCCAGTG 180
QY 181 AATATGGCAACATCAAGATGAGCGCGGTTCCCGAGCAGACTGGTTATCAAGC 240
DB 181 AATATGGCAACATCAAGATGAGCGCGGTTCCCGAGCAGACTGGTTATCAAGC 240
QY 241 CCCCGGCAGAGTTACATTAAGATGAGGTGTACCAACCAACAGGTTAATGGGTCAAGG 300
DB 241 CCCCGGCAGAGTTACATTAAGATGAGGTGTACCAACCAACAGGTTAATGGGTCAAGG 300
QY 301 ATTCACTGATGACTGACGAGCGTGGCAAGGAGGAAATGTTAGCAGTTCAGACAATG 360
DB 301 ATTCACTGATGACTGACGAGCGTGGCAAGGAGGAAATGTTAGCAGTTCAGACAATG 360
QY 361 TTGGATGAATATGGAAGTACATGGAAGAGAGCATATTCGCGCTCCAAATATGACAA 420
DB 361 TTGGATGAATATGGAAGTACATGGAAGAGAGCATATTCGCGCTCCAAATATGACAA 420
QY 421 CCAATGACGAGAGTATTGTCGAGCATCTTACGTTATGAGCAGACACCATGTAC 480
DB 421 CCAATGACGAGAGTATTGTCGAGCATCTTACGTTATGAGCAGACACCATGTAC 480

481 GCAGTGGCTGGAGTGGGAGTGAAGAGTATGTCTTCCAGAGTGGACATCTTGTGT 540
481 GCAGTGGCTGGAGTGGGAGTGAAGAGTATGTCTTCCAGAGTGGACATCTTGTGT 540
541 TCCAGAACATTGATGGGAAAGAGTTGTATAAATACCAAAAGATGACTTCCAGAGACTCA 600
541 TCCAGAACATTGATGGGAAAGAGTTGTATAAATGACCAAAAGATGACTTCCAGAGACTCA 600
601 GCGGAGCTATAACGAGATATCTCTGTGTCACAGTACATCTACCTCAGAGAGA----- 654
601 GCGGAGCTATAACGAGATATCTCTGTGTCACAGTACATCTACCTCAGAGAGACTCTC 660
655 -----
661 TTCCACATTTGACTTCAGATGATGTTGATAAGGCTTACAAAATCTCTCCCGTTAATGC 720
655 -----GAGGAGCCATTTTATTTTCCAAATACATCAGTTTACCAGAG 699
721 ATGCTAGAAAACACAGGAGGAGCCATTTTATTTTCCAAATACATCAGTTTACCAGAG 780
700 CAAACGAAAGATTAACAACAGGCGAGATTACCTTATGAGCAAGCGAGAGATCAGCGT 759
781 CAAACGAAAGATTAACAACAGGCGAGATTACCTTATGAGCAAGCGAGAGATCAGCGT 840
760 GGACGAGTACAGGCCATCCCACTCAGTCAAAAGCTTACCAACCATCATCTTCAACAGTGC 819
841 GGACGAGTACAGGCCATCCCACTCAGTCAAAAGCTTACCAACCATCATCTTCAACAGTGC 900
820 CCAAAACAGAGACCGAGCTCTCAGTTAGATCCCTTATCAGATTTCTTGAGCGACCA 879
901 CCAAAACAGAGACCGAGCTCTCAGTTAGATCCCTTATCAGATTTCTTGAGCGACCA 960
880 GCGGTCTTGCAAAATCCAGGAGTGGGAGATACAGTATGCGAGTTCTTCTGAGGCTTC 939
961 GCGGTCTTGCAAAATCCAGGAGTGGGAGATACAGTATGCGAGTTCTTCTGAGGCTTC 1020
940 TGTGCGAGAGTCCCAACTCCAACTGATCACTGGGAGGACAAATGGGAGTTCAAGA 999
1021 TGTGCGAGAGTCCCAACTCCCACTGATCACTGGGAGGACAAATGGGAGTTCAAGA 1080
1000 TGACAGAGCCCTGATGAAGTGGCTCGGCGTTGGGAGAGAGAGAAAGCAAACTAACATGA 1059
1081 TGACAGAGCCCTGATGAAGTGGCTCGGCGTTGGGAGAGAGAGAAAGCAAACTAACATGA 1140
1060 ACTATGACAAACTCAGCGTGCACCTCGCTACTACTATGACAAATAATTTATGACTAAG 1119
1141 ACTATGACAAACTCAGCGTGCACCTCGCTACTACTATGACAAATAATTTATGACTAAG 1200
1120 TTTCACTGGTAAACGCTATGCCTACAAATTTGATTTTCCAGGAATCGCTCAGGCGCTCCAGC 1179
1201 TTTCACTGGTAAACGCTATGCCTACAAATTTGATTTTCCAGGAATCGCTCAGGCGCTCCAGC 1260
1180 CTCACCGCTCCAGAACTCAGTACAAATACCAATCCGATCCGCTCCGCTACATGAGTTCT 1239
1261 CTCACCGCTCCAGAACTCAGTACAAATACCAATCCGATCCGCTCCGCTACATGAGTTCT 1320
1240 ACCATGCAACACCCCGAGAGATGAATTTGTAGCTTCCCATCCGCTCCGCTTTCGCCGTAA 1299
1321 ACCATGCAACACCCCGAGAGATGAATTTGTAGCTTCCCATCCGCTCCGCTTTCGCCGTAA 1380
1300 CTTCACTCAGCTTTTGTGTCGCCCTAATCCATCTGGAATTCACCAACTGGAGGATCT 1359
1381 CTTCACTCAGCTTTTGTGTCGCCCTAATCCATCTGGAATTCACCAACTGGAGGATCT 1440
1360 ACCCAATATACAGGCTGCCAGCTGCTCATATGCCCTTCCCATCTTGGCACCTACTACTAG 1419
1441 ACCCAATATACAGGCTGCCAGCTGCTCATATGCCCTTCCCATCTTGGCACCTACTACTAG 1500
1420 TGGGAAAGAAAGAGCGCCCAAGAAA 1447
1501 TGGGAAAGAAAGAGCGCCCAAGAAA 1528

```

```

RESULT 4
GGERG
LOCUS      G.gallus ERG mRNA.          1516 bp      mRNA      linear      VRT 27-APR-1995
DEFINITION X77159
ACCESSION  X77159
VERSION     X77159.1 GI:790439
KEYWORDS   Gallus gallus (chicken)
SOURCE     Gallus gallus
ORGANISM   Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.

REFERENCE
1
AUTHORS   Dordain,P., Dewitte,F., Desbiens,X., Stehelin,D. and
            Dutercq-Coquillaud,M.
TITLE     Mesodermal expression of the chicken erg gene associated with
            precartilaginous condensation and cartilage differentiation
JOURNAL   Mech. Dev. 50 (1), 17-28 (1995)
MEDLINE   95329425
PUBMED    7605748
REFERENCE
2 (bases 1 to 1516)
Dutercq-Coquillaud,M.
Direct Submission
Submitted (17-JAN-1994) M. Dutercq-Coquillaud, CNRS UA 1160,
Oncologie Moleculaire, Institut Pasteur, 1 rue Calmette, 59019
Lille, FRANCE
FEATURES
            Location/Qualifiers
            1..1516
               /organism="Gallus gallus"
               /mol_type="mRNA"
               /db_xref="taxon:9031"
               /tissue_type="spleen"
               /clone_lib="lambda gt10"
               /dev_stage="adult"
            1..1516
               /genes="ERG"
               /gene="ERG"
               /codon_start=1
               /protein_id="CAA54404.1"
               /db_xref="GI:790440"
               /db_xref="GOA:Q90837"
               /db_xref="SWISS-PROT:Q90837"
               /translation="MASTIKALSVSEDSQSLFECAYGSPHLAKTEMTASSSEYQGT
               SKMSRPVQDMLSQSPARVITIKMECNQVNGSRNSPDDCSVAKGKMVSSSDNVGM
               NYGSMYBEXHIPPNNPTNRRVIVPADPTLWSTDHVRQWLEWAVEYGLPVDILLF
               ONTDGKELCKWTDDFORLTPSYNADILLSHLYRETLPPLHSDVDVKALQNSPRL
               MHRNFGGATFIPNYSVFEATQRTITRPDLPEQARRASANTSHSPQSKATQPSL
               STVPTEDQRPDQVQILGPTSSRLNPGSGQIQWQELLELLSDSSNSNCITWEGT
               NGEFKMTDPEVARRWGERKSKPNMYDKLSRALRYDDKINIMTKVHGKRYAKPDPFH
               GIAALQHPHPESMYKPSDLPYNGSYFAHPQKMNFAVPHPPALPVTSSSFFAAPNP
               YMNSPTGGIYVENTRLPAAHMPSHLGTYY"

Query Match      92.9%; Score 1344; DB 5; Length 1516;
Best Local Similarity 94.7%; Pred. No. 0;
Matches 1435; Conservative 0; Mismatches 0; Indels 81; Gaps 1;

QY      1  GAATTCGCGAAGCAATAATTATTATTAGCAATTATTAGCGATCAATAATCTTGATCACA 60
DB      1  GAATTCGCGAAGCAATAATTATTATTAGCAATTATTAGCGATCAATAATCTTGATCACA 60

QY      61  TTATGGCAAGCACTATTAAAGAGCACTATCAGTGGTGAAGACCACTGCTTGTGTTG 120
DB      61  TTATGGCAAGCACTATTAAAGAGCACTATTATCAGTGGTGAAGACCACTGCTTGTGTTG 120

QY      121  AGTGTGCTACGGATCGCCCACTTCCAAAGACAGAAATGACAGCTCTCTTCCAGTG 180
DB      121  AGTGTGCTACGGATCGCCCACTTCCAAAGACAGAAATGACAGCTCTCTTCCAGTG 180

QY      181  AATATGGGCAAAACATCAAAAGATAGCCGCGTTCGCCAGAGGAGCTGTTATCA 240
DB      181  AATATGGGCAAAACATCAAAAGATAGCCGCGTTCGCCAGAGGAGCTGTTATCA 240

```

## ORIGIN

```

241  CCCCGCCAGAGTTACCATTAAGATGGAGTGTAAACCCAAAACAGGTTAATGGGTCAAGA 300
DB      CCCCGCCAGAGTTACCATTAAGATGGAGTGTAAACCCAAAACAGGTTAATGGGTCAAGA 300

301  ATTACCTGATGACTGCGAGCGTGGCAAAAGAGGAGGAAATGGTTAGCAGTTTCAGACAATG 360
DB      ATTACCTGATGACTGCGAGCGTGGCAAAAGAGGAGGAAATGGTTAGCAGTTTCAGACAATG 360

361  TTGGGATGAACCTATGGAAGCTACATGGGAAGAGACATATTCGGCTCCAAATATGACAA 420
DB      TTGGGATGAACCTATGGAAGCTACATGGGAAGAGACATATTCGGCTCCAAATATGACAA 420

421  CCAATGAACGAAGAGTTATTGTGCCAGCAGATCTCTACCTATTGGAGCAGACAGCAATGTAC 480
DB      CCAATGAACGAAGAGTTATTGTGCCAGCAGATCTCTACCTATTGGAGCAGACAGCAATGTAC 480

481  GGCAGTGGCTGGAGTGGCAGTGAAGGAGTATGGTCTTCAGACGTCGACATCTTGTGT 540
DB      GGCAGTGGCTGGAGTGGCAGTGAAGGAGTATGGTCTTCAGACGTCGACATCTTGTGT 540

541  TCCAGAACATTGATGGGAAAGAGTTGTGTAAATGACCAAAAGAGTCTTCAGAGACTCA 600
DB      TCCAGAACATTGATGGGAAAGAGTTGTGTAAATGACCAAAAGAGTCTTCAGAGACTCA 600

601  CGCCGAGCTATAACGCACATATCTCTCTGTACACCTACCTACCTACAGAGACTCCTC 660
DB      CGCCGAGCTATAACGCACATATCTCTCTGTACACCTACCTACCTACAGAGACTCCTC 660

655  -----
DB      -----

661  TTCCACATTTCAGATTCAGATGATGTTGATAAGGCGCTTACAAAACCTCCACGGTTAATGC 720
DB      TTCCACATTTCAGATTCAGATGATGTTGATAAGGCGCTTACAAAACCTCCACGGTTAATGC 720

655  -----
DB      -----

721  ATGCTAGAAACACAGGAGGAGGCCACTTTTATTTTCCAAATACATCAGTTTACCCAGAAG 780
DB      ATGCTAGAAACACAGGAGGAGGCCACTTTTATTTTCCAAATACATCAGTTTACCCAGAAG 780

700  CAAAGCAAGAAATAACAACAAGGCCAGATTACCTTATGAGCAAGCGAGGAGATCAGCGT 759
DB      CAAAGCAAGAAATAACAACAAGGCCAGATTACCTTATGAGCAAGCGAGGAGATCAGCGT 759

781  CAAAGCAAGAAATAACAACAAGGCCAGATTACCTTATGAGCAAGCGAGGAGATCAGCGT 840
DB      CAAAGCAAGAAATAACAACAAGGCCAGATTACCTTATGAGCAAGCGAGGAGATCAGCGT 840

760  GGAAGAGTCAAGCCATCCCACTCAGTCAAAAGCTACCCAAACCATATCTTCAACAGTGC 819
DB      GGAAGAGTCAAGCCATCCCACTCAGTCAAAAGCTACCCAAACCATATCTTCAACAGTGC 819

841  GGACGAGTCACAGCCATCCCACTCAGTCAAAAGCTACCCAAACCATATCTTCAACAGTGC 900
DB      GGACGAGTCACAGCCATCCCACTCAGTCAAAAGCTACCCAAACCATATCTTCAACAGTGC 900

820  CCAAAACAGAAAGACAGAGTCTCCTCAGTTAGATCCTTATCAGATTTCTTGACCGACCA 879
DB      CCAAAACAGAAAGACAGAGTCTCCTCAGTTAGATCCTTATCAGATTTCTTGACCGACCA 879

901  CCAAAACAGAAAGACAGAGTCTCCTCAGTTAGATCCTTATCAGATTTCTTGACCGACCA 960
DB      CCAAAACAGAAAGACAGAGTCTCCTCAGTTAGATCCTTATCAGATTTCTTGACCGACCA 960

880  GCCGTCTTTGCAAAATCCAGGAGTGGCGAGATACAGCTATGCGAGTTCCTATGGAGCTTC 939
DB      GCCGTCTTTGCAAAATCCAGGAGTGGCGAGATACAGCTATGCGAGTTCCTATGGAGCTTC 939

940  TGTCCGACAGCTCCAACTCCAACTGCATCAGCTGGGAGGGCACAATGGGAGTTCCAGA 999
DB      TGTCCGACAGCTCCAACTCCAACTGCATCAGCTGGGAGGGCACAATGGGAGTTCCAGA 999

1021  TGTCCGACAGCTCCAACTCCAACTGCATCAGCTGGGAGGGCACAATGGGAGTTCCAGA 1080
DB      TGTCCGACAGCTCCAACTCCAACTGCATCAGCTGGGAGGGCACAATGGGAGTTCCAGA 1080

1000  TGACAGACCCCTGATGAGTGGCTCGCGCTTGGGAGAGAGGAAAGAAACCTTAACATGA 1059
DB      TGACAGACCCCTGATGAGTGGCTCGCGCTTGGGAGAGAGGAAAGAAAGAAACCTTAACATGA 1059

1081  TGACAGACCCCTGATGAGTGGCTCGCGCTTGGGAGAGAGGAAAGAAAGAAACCTTAACATGA 1140
DB      TGACAGACCCCTGATGAGTGGCTCGCGCTTGGGAGAGAGGAAAGAAAGAAACCTTAACATGA 1140

1060  ACTATGACAAACTCAGCGCTGCCTTCCCTACTACTATGACAAATAATATTATGCTAAG 1119
DB      ACTATGACAAACTCAGCGCTGCCTTCCCTACTACTATGACAAATAATATTATGCTAAG 1119

1141  ACTATGACAAACTCAGCGCTGCCTTCCCTACTACTATGACAAATAATATTATGCTAAG 1200
DB      ACTATGACAAACTCAGCGCTGCCTTCCCTACTACTATGACAAATAATATTATGCTAAG 1200

1120  TTCTGTTAAACGCTATGCTTCAAAATTTGATTTCCAGGAATCGCTCAGGCCCTCCAGC 1179
DB      TTCTGTTAAACGCTATGCTTCAAAATTTGATTTCCAGGAATCGCTCAGGCCCTCCAGC 1179

1201  TTCTGTTAAACGCTATGCTTCAAAATTTGATTTCCAGGAATCGCTCAGGCCCTCCAGC 1260
DB      TTCTGTTAAACGCTATGCTTCAAAATTTGATTTCCAGGAATCGCTCAGGCCCTCCAGC 1260

1180  CTCACCTCCAGAAATCATCCATGTACAAATACCCATCAGCTCCCTCCCTACAGTCTCT 1239
DB      CTCACCTCCAGAAATCATCCATGTACAAATACCCATCAGCTCCCTCCCTACAGTCTCT 1239

1261  CTCACCTCCAGAAATCATCCATGTACAAATACCCATCAGCTCCCTCCCTACAGTCTCT 1320
DB      CTCACCTCCAGAAATCATCCATGTACAAATACCCATCAGCTCCCTCCCTACAGTCTCT 1320

```

```
QY 1240 ACCATGACACCCCGACAGAGTGAACCTTTGTAGCTCCCATCCCTGCTTGCCTGTA 1299
Db 1321 ACCATGACACCCCGACAGAGTGAACCTTTGTAGCTCCCATCCCTGCTTGCCTGTA 1380
QY 1300 CCTCATCAGCTTTTGTGCTCCCTTAATCATCTAGTGAATTCACCACTGGAGGATCT 1359
Db 1381 CCTCATCAGCTTTTGTGCTCCCTTAATCATCTAGTGAATTCACCACTGGAGGATCT 1440
QY 1360 ACCCAATACAGAGCTGCAGCTGCTCATATGCTTCCCATCTGGCACTACTACTAAG 1419
Db 1441 ACCCAATACAGAGCTGCCAGTCTCATATGCTTCCCATCTGGCACTACTACTAAG 1500
QY 1420 TGGGAAAGAAAGAAA 1435
Db 1501 TGGGAAAGAAAGAAA 1516

RESULT 5
AX657228
LOCUS AX657228 1509 bp DNA linear PAT 22-MAR-2003
DEFINITION Sequence 1 from Patent WO03000724.
ACCESSION AX657228
VERSION AX657228.1 GI:29160006
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Luyten,F., de Bari,C. and Dell'Accio,F.
TITLE Polynucleotide sequences and vectors useful for the prevention or
JOURNAL treatment of bone- or cartilage-related disorders
PATENT: WO 03000724-A 1 03-JAN-2003;
Tigenix N.V. (BE)
FEATURES
Location/Qualifiers
source 1..1509
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
67..1446
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD79711.1"
/db_xref="GI:29160007"
/db_xref="REMBL:CAD79711"
/translation="MIOTVDPAAHLKEALSVVSEDSLPFECAYGTPLAKTMTBASS
SSDYQTSKMSRPVQDNLSQPPARTVIMKCNPSQVNGSRNSPSECSVAKGKMGV
SPDVTGMNYSIMEEKMPFPNMTNERRVIVPADPLMSTIDHVRWLEWAVKEYGLP
DVNILLFQNDIGKELCKMTXDDFQRLTPSYNADILLSHLYLHREGAAFIFFNTSYP
EATQRTITREDLPYEPERRSAWTHGHTPQSKAQPSPSTVPTKEDQRPOLDPYQIL
GPTSSRLANFGSGQIQWLQFLLELLSDSNSSCITWEGTNGEFKMTDPOEVARWGER
KSKPNMYDKLSRALRYDYDKNTIMTKVHGKRYAYKPDFHGIAQLQPHPPESLYKYP
SDLPYMGSYHAHPQKMFVAPHPALPVTSSSFFAAPFPVWNSPTGIVPNTRLPTSH
MPSHLGYTY"
CDS
Query Match 68.7%; Score 994.4; DB 6; Length 1509;
Best Local Similarity 84.1%; Pred. No. 7.9e-290;
Matches 1134; Conservative 0; Mismatches 211; Indels 3; Gaps 1;

QY 74 TATTAAAGAACATTATCAGTGGTGAAGTGAAGACCGCTCTCTTCCAGTGAATATGGCAAC 133
Db 99 TATCAAGAGACCTTATCAGTTGTGAGTGAGGACGACGTCTGTGTGAGTGTGCTTACGG 158
QY 134 ATCGCCCCACCTTGCAAGACAGAAATGACAGCGCTCTCTTCCAGTGAATATGGCAAC 193
Db 159 AACGCCACCTGGCTAAGACAGAGATGACCGCGTCTCTCTCCAGCGACTATGGACAGAC 218
QY 194 ATCAAGATAGCCCGCGTTCCTCCAGCAGGACTGGTTATCAGAGCCCCCGGCAGAGT 253
Db 219 TTCCAAGATAGCCACCGCTCCCTCAGCAGGATTGGCTGTCTCAACCCCCCAGCCAGGT 278
QY 254 TACCATTAAAGATGAGTGTAAACCAACACAGGTAAATGGGTCAAGGAATTCACCTGATGA 313
```

```
Db 279 CACCATCAAAATGGAAATGTAACCTTAGCCAGGTGAATGGCTCAAGAACTCTCCTGATGA 338
QY 314 CTGAGCGCTGGCAAAAGAGGGAAAATGGTTAGCAGTTTCAGCAATGTTGGGATGAACATA 373
Db 339 ATGCGAGTGTGGCCAAAGCGGGAAGATGGTGGGAGAGCCACACACCCTGTGGATGAACATA 398
QY 374 TGGAAAGCTACATGGAAGAGAAAGCATATTCGCGCTCCAAATATGACAACTAATGAACGAAG 433
Db 399 CGGCGAGCTACATGGAAGAGAAAGCATATGCGCAACCCCAACATGACCAAGAGCGGAG 458
QY 434 AGTTATTGCGCAGAGATCTTACGTTATGAGACACAGACCATGTACGGCAGTGGCTGGA 493
Db 459 AGTTATGTCGCGCAGAGATCTTACGCTATGAGTACAGACCATGTGCGGCGAGTGGCTGGA 518
QY 494 GTGGCGAGTGAAGGAGTATGGTCTTCCAGACGTCAGACATCTTGTGTTCCAGAAACATGA 553
Db 519 GTGGCGGTGAAAGAAATATGGCTTCCAGACGTCACATCTTGTATTCCAGAACATCGA 578
QY 554 TGGGAAAGAGTTGTGTAAATGACCAAGATGACTTCCAGAGACTCAGCGGAGGTATAA 613
Db 579 TGGGAAAGAACTGTGCAAGATGACCAAGGAGACTTCCAGAGGCTCACCCCGAGTACAA 638
QY 614 CGCAGATATCTCTCTGTCACACCTACACTACTCAGAGAGAGAGGACACTTTTATTTT 673
Db 639 CGCCGACATCTCTCTCTCAGATCTCCACTACTCAGAGAGAGGAGTGGAGCTTTTATTTT 698
QY 674 TCCAAATACATCAGTTTACCAGAAAGCAACGAAAGATAAACAAGCGCCAGATTTACC 733
Db 699 CCAAAATATCTTCAGTATATCTGAAAGCTACGCAAAAGATTAACAATAGGCCAGATTTACC 758
QY 734 TTATGACGCAAGGAGGAGATCAGGTGACGAGTCACAGGCTACAGCCATCCCA---CTCAGTCAA 790
Db 759 ATATGAGCCCCCAGGAGATCAGCCTTGAGCGGTGACGGCCACCCAGCCCGAGTCCGAA 818
QY 791 AGCTACCAACCATCATCTTCAACAGTGCCTCAAAAGACAGAAAGACAGCGCTCTCAGTTAGA 850
Db 819 AGCTGCTCAACCATCTCTTCCACAGTGCCTCAAAAGTGAAGACAGCGCTCTCAGTTAGA 878
QY 851 TCCTTATCAGATCTTGGACCGACGAGCGGTGTCGAAATCCAGGAGTGGGAGAT 910
Db 879 TCCTTATCAGATCTTGGACCAACAAAGTAGCGCGCTTGCAAAATCCAGGAGTGGCGAGAT 938
QY 911 ACAGTATGCGAGTTCTTACTTGGAGCTTCTGTCGACAGCTCCAACTCCAACTCAGCTCAC 970
Db 939 CCAGCTTTGGCAGTTCTCTCTGAGGCTCTCTGTCGACAGCTCCAACTCAGCTCAC 998
QY 971 CTGGAGGCGCAAAATGGGAGTTCAGATGACAGACCTGATGAGTGGCTCGCGCTG 1030
Db 999 CTGGGAAGGCGCAACCGGGAGTTCAAGATGACGATCCCGACGAGGTGGCGCGCGCTG 1058
QY 1031 GGGAGAGAGGAAAGCAACCTTAACATGAATATGACAAACTCAGCGCTGCACTTCGGTA 1090
Db 1059 GGGAGAGCGGAAGAGCAAAACCAACATGACTAGATAAGCTCAGCGCGCTCCGTTA 1118
QY 1091 CTACTATGACAAAATATTTATGACTAAAGTTTCATGTTAAAGCTATGCTACAAATTTGA 1150
Db 1119 CTACTATGACAAAGACATCATGACCAAGGTCATGGGAAGCGCTACGCTCAAGTTGGA 1178
QY 1151 TTTCCAGGAACTCGCTCAGGCGCTCCAGCTCACCTCCAGAACTCATCATGTACAAATA 1210
Db 1179 CTTCACGGGATCGCCGAGGCGCTCCAGCCCAACCCCGGAGTCACTCTGTACAAGTA 1238
QY 1211 CCGATCAGACCTTCCCTTACATGAGTTCTTACATGCAACCCCGCAGAGATGAATTTGT 1270
Db 1239 CCGCTCAGACCTTCCCGTACATGGGCTCTCTATCAGCCCAACCCCAAGAGATGAATTTGT 1298
QY 1271 AGCTCCCATCCCTGCTTTGCCGTAACCTCATCAGCTTTTGTGCTCCCTTAATCC 1330
Db 1299 GGGCGCCCAACCTCAGCGCTCCCGTGACATCTTCCAGTTTTTGTGCTGCCCAACCC 1358
QY 1331 ATACTGGAATTCACCAACTGGAGGATCTAOCCTAATACAGGCTGCGCAGCTGTCTATAT 1390
```

```

Db      1359 ATACTGGAATTCACAACTGGGGGTATATATACCCACACTAGGCTCCCAACAGCCATAT 1418
Qy      1391 GCCTTCCACTCTTGGCACTACTACTAA 1418
Db      1419 GCCTTCACTCTGGCACTACTACTAA 1446

RESULT 6
LOCUS   BC040168
DEFINITION Homo sapiens v-ets erythroblastosis virus E26 oncogene like
          (avian) transcript variant 1, mRNA (CDNA clone MGC:49020
          IMAGE:6052140), complete cds.
ACCESSION BC040168
VERSION   BC040168.1 GI:25304065
KEYWORDS MGC.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2016)
AUTHORS  Krausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
          Klausner,R.D., Collins,F.S., Wagner,L., Shennen,C.M., Schuler,G.D.,
          Altshuler,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
          Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
          Diatchenko,L., Narusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
          Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
          Scheetz,T.E., Brownstein,M.J., Udell,T.B., Toshiyuki,S.,
          Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
          Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
          McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
          Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
          Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
          Sanchez,J., Helton,E., Kettman,M., Madan,A., Rodrigues,S.,
          Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
          Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
          Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
          Butlerfield,Y.S., Krzywinski,M.I., Skalska,U., Smalish,D.E.,
          Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
          Generation and initial analysis of more than 15,000 full-length
          human and mouse cDNA sequences
          Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
          22388257
PUBMED  12477932
REFERENCE 2 (bases 1 to 2016)
AUTHORS  Krausberg,R.
TITLE    Direct Submission
JOURNAL  Submitted (22-NOV-2002) National Institutes of Health, Mammalian
          Gene Collection (MGC), Cancer Genomics Office, National Cancer
          Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
          USA
REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT  Contact: MGC help desk
          Email: cgapbs-remail.nih.gov
          Tissue Procurement: ATCC
          cDNA Library Preparation: Life Technologies, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Institute for Systems Biology
          http://www.systemsbio.org
          contact: amadan@systemsbiology.org
          Anup Madan, Jessica Fahney, Erin Helton, Mark Kettelman, Anuradha
          Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 84 Row: 0 Column: 9.
          Location/Qualifiers
          1..2016
          /organism="Homo sapiens"
          /mol_type="mRNA"
          /db_xref="taxon:9606"
          /clones="MGC:49020 IMAGE:6052140"
          /tissue_type="Testis, embryonal carcinoma"

FEATURES
Source

```

```

/clone_lib="NIH_MGC_92"
/lab_host="DH10B"
/notes="Vector: pCWV-SPORT6"
1..2016
/genes="ERG"
/db_xref="LocusID:2078"
/db_xref="MIM:165080"
104..1543
/codon_start=1
/product="v-ets erythroblastosis virus E26 oncogene like,
isoform 1"
/protein_id="AAH40168.1"
/db_xref="GI:25304065"
/translation="MASTIKKALSVSDQSLFECAYGTPLAKTMTASSSSDVGQT
SKMSPRVQDQWLSQPARVTIKMPCNSQVNGSRNSPDECVAKGKGVGSPDTVGM
NYGSMKEMHPPNMTNERRVTPADPTLWSTDHVQLEWAKVGLDPLVNILLF
QNIDGKELCKMTKDDFORLTYNADILLSHLHRETPHLTDSDDVDKALQNSPRL
MHARTGGAARIFPNTSVYPTAOTITRPDLVPYPPERSAWTGHGHTPKOSKAQPS
FSTVFKISDQRPLQDPYQILGFTSRLANPSSGGIQLWQFLELLSDSSNSCITWEG
TNGEFMTDTPDEVARWGERKSNMNYDKUSRALRYIDKNIMTKVHKRYAYKFD
HGIAALQHPHPESSLYKYPDSLPMYGSYHAHPQDNFVAPHPPALPVTSSSFFAAPN
PYMNSPTGGIYPTNRLPTSHMPSHLGTY"
446..694
/notes="SAM_PNT; Region: Sterile alpha motif (SAM)/Pointed
domain"
/db_xref="CDD:pfam02198"
1031..1282
/notes="Ets; Region: Bts-domain"
/db_xref="CDD:pfam00178"

ORIGIN
Query Match 65.9%; Score 953.4; DB 9; Length 2016;
Best Local Similarity 79.6%; Pred. No. 2.2e-277; Indels 84; Gaps 2;
Matches 1190; Conservative 0; Mismatches 221;
Qy 8 GCGAAGCAATATATTATTAGCAATATTAGCGATCAATATCTTGATCATATATGCG 67
Db 49 GAGAAGCAATATATTATTAGCAATATTAGCGATCAATATCTTGATCATATATGCG 108
Qy 68 AAGCAATATTAGGAAGCAATATCTAGTGTGATGAGCAAGCAAGCAAGTGTGAGTGTG 127
Db 109 CAGCAATATTAGGAAGCAATATCTAGTGTGATGAGCAAGCAAGTGTGAGTGTG 168
Qy 128 CTACGATGCGCCACCTTGCAGACAGAAATGACAGCTCTCTCTCCAGTGAATATGG 187
Db 169 CTACGAGAGCCACACCTTGGCTAGACAGAGATGACCGGCTCTCTCCAGCACTATGG 228
Qy 188 GCAACATCAAGATGAGCGCGCGGTTCGCCAGAGGACTGGTTATCACAGCCCCCGGC 247
Db 229 ACAGACTTCCAGATGAGCGCGGTTCGCCAGAGGACTGGTTATCACAGCCCCCGGC 288
Qy 248 CAGAGTTACCATTAAGATGGAGTGAACCCAAACAGGTTAATGGTCAAGGAATTCACC 307
Db 289 CAGGTCACCATCAAAATGGATGTAACTTAGCAGGTGATGGCTCAAGGAATCTCC 348
Qy 308 TGATCACTCAGCGTGGGCAAGAGGAGAAATGGTTAGCAGTTTCAAGCAATTTGGGAT 367
Db 349 TGATCAATGTCAGTGTGGCCAAAGGCGGGAAGATGTGTGGCGCAGCCAGACACCGTTGGGAT 408
Qy 368 GAACCTAGAGCTACATGGAGAGAGCAATATTCGGCTCCCAATATGACAAACAATGA 427
Db 409 GAACCTAGAGCTACATGGAGAGAGCAATATTCGGCTCCCAATATGACAAACAATGA 468
Qy 428 ACGAAGAGTTATTGTGCCAGCAGATCTTACGTTATGGAGCAGACAGACCATGTAGCGAGTG 487
Db 469 GCGCAGAGTTATCGTGTGCCAGCAGATCTTACGTTATGGAGTACAGACCATGTAGCGAGTG 528
Qy 488 GCTGAGTGGCGAGTGAAGGAGTATGGTCTTCCAGACGTTGACATCTTGTGTTCCAGAA 547
Db 529 GCTGAGTGGCGGTTGAAAGAAATATGGCTTCCAGACGTTCAACATCTTGTATTCCAGAA 588
Qy 548 CATTGATGGGAAAGAGTTGTGTAAATGACCAAGATGACTTCCAGAGACTCACGCCGAG 607

```

```

Db      589 CATGATGGGAAGAACTGTCTCAAGATGACCAAGAGACGATTCAGAGGCTCACCCCCAG 648
QY      608 CTATAACCCAGATATCTCTCTGTGTACACCTACACTACTCTCAGAGAGA-----654
Db      649 CTACAATCCGACATCTCTCTCTCACATCTCCACTAGCTCAGAGAGACTCTCTTCCACA 708
QY      655 -----654
Db      709 TTTGACTTCAGATGTTGTATAAGCCTTACAAAATCTCTCCAGGTTAATGATGCTAG 768
QY      655 -----GAGAGCCACTTTTATTTTCCAAATACATCACTTTACCCAGAACGCA 706
Db      769 AAACACAGGGGTGCAGCTTTTATTTTCCCAATACTTCACTATATCTCTGAAGTACGCA 828
QY      707 AAGATATACACACAGGCCAGATTTACCTTATGACCAAGCGAGGAGATCAGCGTGGACG 766
Db      829 AAGAAATTAACAATAGGCCAGATTTACCAATATGAGCCCCCAGGAGATCAGCGTGGACCG 888
QY      767 TCACAGCCATCCCA---CTCAGTCAAAAGCTTACCAACCATCATCTTCAACAGTGCCTCAA 823
Db      889 TCAGGCCACCCACGCCCGCAGTCGAAAGCTGTCTCAACATCTCTTCCACAGTGCCTCAA 948
QY      824 AACAGAACACAGCGTCTCTAGTTAGATCTTATCAGATCTTTGGACCGACAGCAGCG 883
Db      949 AACTGAACACAGCGTCTCTAGTTAGATCTTATCAGATTTCTTGGACCAACAAGTAGCG 1008
QY      884 TCTTCAAAATCCAGGAGTGGCAGATACAGCTATGCGAGTTCCTACTTGGAGCTTCTGTG 943
Db      1009 CTTTCAAAATCCAGGAGTGGCAGATCCAGCTTTGGCAGTTCTCTCTGGAGCTCTGTG 1068
QY      944 GGACAGTCTCAACTCACTGATCACTCTGGAGGGGACAAATGGGAGTTCAGATGAC 1003
Db      1069 GGACAGTCTCAACTCACTGATCACTCTGGAGGGGACCAACCGGGAGTTCAGATGAC 1128
QY      1004 AGACCTCATGAAGTGGTTCGGCGTTGGGAGAGAGAAAGCAAACTAACATGAACCTA 1063
Db      1129 GGATCCCGACGAGTGGCGCGCTGGGAGAGCGGAGAGCAAAACCAACATGAACCTA 1188
QY      1064 TGACAAACTCAGCGTGCATCTGCTACTACTATGACAAATAATTAATGACTAAAGTTCA 1123
Db      1189 CGATAAGCTCAGCGCGCGCTCCGTTACTATGACAAAGAAACATCATGACCAAGTCCA 1248
QY      1124 TGGTAAACGCTATGCTCAAAATTTGATTTCCAGGATCGCTCAGGCGCTCCAGCCTCA 1183
Db      1249 TGGGAAGCGCTACGCGCTCAAGTTGAGATTCAGATTCACGCGGATGCGCCAGGCGCTCCAGCCCCA 1308
QY      1184 CCCTCCAGATCATCCATGTACAAATACCCATCAGACCTCCCTTACATGAGTTCTTACCA 1243
Db      1309 CCCCCCGGAGTCACTCTGTACAAGTACCCCTCAGACCTCCGTTACATGGCTCTATCA 1368
QY      1244 TGCACACCCCGACAGATGAATTTGAGCTCCCATCCCTCCCTGCTTGGCCGTTAACTC 1303
Db      1369 CGCCACCCACAGAGATGAATTTGTTGGCGCCCAACCCCTCAGCCCTCCCGGTGACATC 1428
QY      1304 ATCCAGCTTTTGTGTCGCCCTTAATCCATCTGGAATTCACCAACTCGAGGCACTTACCC 1363
Db      1429 TTCCAGTTTGTGTCGCCCAACCCCATCTGGAATTCACCAACTCGGGGTATATACC 1488
QY      1364 CAATACAGGCTGCCAGTGTCTATGCTTCCATCTTGGACCTTACTACTAA 1418
Db      1489 CAACACTAGGCTCCCGACCGCATATGCTTCTCATCTGGGCACTTACTACTAA 1543

RESULT 7
LOCUS   AB031088
DEFINITION Rattus norvegicus VESPI4 mRNA for vascular endothelial cell specific protein 14, complete cds.
ACCESSION AB031088
VERSION   AB031088.1
KEYWORDS vascular endothelial cell specific protein 14.
SOURCE   Rattus norvegicus (Norway rat)

```

```

ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 1808)
AUTHORS Aoki,T., Toyoda,H., Nishimoto,S., Tawara,J., Ukai,Y. and
Komuraasaki,T.
TITLE Identification of VESPI4, a vascular endothelial cell specific
protein
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1808)
AUTHORS Aoki,T., Toyoda,H., Nishimoto,S., Tawara,J., Ukai,Y. and
Komuraasaki,T.
TITLE Direct Submission
JOURNAL Submitted (12-AUG-1999) Tomomi Aoki, Taisho Pharmaceutical Co.,
Ltd., Medicinal Research Laboratories, Molecular Biology Lab.,
1-403, Yoshino-cho, Omiya, Saitama 3308530, Japan
(E-mail:sl7704@ccm.taisho.co.jp, Tel:048-663-1111,
Fax:048-652-7254)
FEATURES
source
1..1808
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/tissue_type="liver"
1..1808
/gene="VESPI4"
63..1430
/gene="VESPI4"
/note="similar to human erg1"
/codon_start=1
/product="vascular endothelial cell specific protein 14"
/protein_id="BAB62744.1"
/db_xref="GI:15128489"
/translation="MASTIKELASVSKDQSLFECAYGTPHLAKTMTASSSDYQGT
SKMSRVQQQWLVSOPPARVTIKMNCNQVNGSRNSPDECSVTGKXVGSPTDVG
SYGSRVSKHMPNMTNERRVTPADPTLWSTDHVQWLEWAVEYGLLDVDVLLF
QNIDKELCKMTKDDFORLTPSYADILLHLVRELPDLTSDDDKALONSPRL
MEARDLTPFRSRSTWGHSHPTPQSKAAQSPSTVKEDQRPQDPQLPILPST
SSLANPGSGQQLWQFLLELSDSNCSNCTWEGTNGFBKMTDDEAVRWGRSKP
NNNYDKLRLRYYYDKNIMTKVHGKRYAYKPFHGIQAOLQHPHPSLSLYKPSDLP
YMGSYTHPQKMFVAPHPALPVTSSSPFATPNFYNSPTGGIYPTNRLPASHMPH
LQTY"
Query Match 64.7%; Score 936; DB 10; Length 1808;
Best Local Similarity 80.2%; Pred. No. 4.2e-272;
Matches 1140; Conservative 0; Mismatches 270; Indels 12; Gaps 3;
QY 8 GCGAACGATAATTATTATTAGCAATTTATTAGCGATCAATAATCTTGATCAATTATGCG 67
Db 8 GAGAGCAATAATTATTATTAACTTTATTACGATCAATAAATTTGATTGATTATGCG 67
QY 68 AAGCACTAATAGGAAGCATTTATCAGTGTGAGTGAAGACCACTCTCTCTTCAGTGAATATG 127
Db 68 CAGCACTAATAGGAAGCGCTTGTCTAGTTGTGAGCAAGACCACTCACTATTGAGTGTGC 127
QY 128 CTACGGATCGCCCCACCTTGCAAGACAGAAATGACAGCTCTCTCTTCAGTGAATATG 187
Db 128 CTACGGAAGCCACACCTGCTGCTAAGACAGAAATGACCGATCTCTCTCCAGTACTATG 187
QY 188 GCAACATCAAGATGAGCCCGCGGTTCCCGCAGGAGCTGTTATCATCAGCCCCCGC 247
Db 188 CCAGACATCAAGATGAGTCCCGAGAGTCCCAACAGGACCTGGCTGTCTCAGCCCCCAGC 247
QY 248 CAGAGTTACCATTAAGATGAGTGAACCCCAACAGGTTAATGGTCAAGGAATTCAC 307
Db 248 CAGGTCACCATCAAGATGAGTGAACCCCTAGCAGGTTGACGTTCCAGGACTCAC 307
QY 308 TGATGACTCCAGCGTGGCAAAAGGAGGAATATGTTAGCACTTCAGACAATGTTGGAT 367
Db 308 TGACGAATGAGTGTGACCAAAAGTGGGAAGATGTTGGGACGCCCTGACTACTGTGGAA 367

```

Qy	368	GA	CTATGGAAGCTA	CA	TGGAAGAGAGCATATTC	CGCCTCCAAATATGACACCAATGA	427
Db	368	GAG	CTACGGCAGTTAC	ATGAGAGAGAGACAT	TCGCGCCCCCAATATGACACGCAATGA	427	
Qy	428	ACG	AAAGAGTTATT	TGTGCGCAGCATCT	CAGTATGAGACACAGACCATGACGCGAGTG	487	
Db	428	ACG	CAGAGTGAT	CGTCCCTCGAGATCT	CTCTGTGAGACACAGACCATGCTCGGCGAGTG	487	
Qy	488	GC	TGGAGTGGCGAGT	GAAGAGTATG	CTTCAGAGCGTGGACATCTTGTTGTTCCAGAA	547	
Db	488	GC	TGGAGTGGCGAGT	GAAGAAATACGCGCTCCT	TAGAGCTGGAGCTCTTATTTATTTCCAGAA	547	
Qy	548	CA	TGTATGGAAAGAGT	TGTGTAATGACCAAGATGACT	TCCAGAGACTCAACGCGAG	607	
Db	548	TA	TGACGCGAAGGAG	CTGTGCAAGATGACAAAGATGACT	TCCAGAGGCTCACTCCGAG	607	
Qy	608	CT	ATACGAGATAT	CTCTCTGTGCACACTAC	CTACTACTCAGAGACTCCCTTCCACA	662	
Db	608	CT	CAATGCGCGA	TTCTTCTCACATCTCCACT	TCTTCTCAGAGACTCCCTTCCACA	667	
Qy	663	ACT	TTTTATTTTTCCAA	TACATCAGTTTATCCAGAA	----GCAACGCAAGAAATAACAAC	718	
Db	668	TC	TGACATCCGACG	AGCTTGATAAGGCTTTACAAA	ACTCTCCACGGTTAATGATGCTAG	727	
Qy	719	AAG	CCAGATTTACCTT	TATGACGAGGAGAGAT	CAGCGTGCAGCGTGCACAGCATCC	778	
Db	728	AA	CACAGATTTACCTT	TATGAGCTCCGAGAGATCA	CACTGACCGGCACACGACACC	787	
Qy	779	CA	---CTCAGTCAAA	AGCTACCCAAACCATCATCTT	CAACAGTGGCCCAAAACAGAGACCA	835	
Db	788	CAC	CCCTCAGTCAAA	AGCTGCCAGCCATCTCCT	TCCAGAGTCCCCAAAACCTGAAGACCA	847	
Qy	836	GC	GTCTCAGTATAG	TCCTTATCAGATCTTGACCGACGACGAGCGTCTTGC	AAATCC	895	
Db	848	GC	GTCTCAGTATAG	TCCTTACAGATCTTGGGACCCAC	CAGTAGTCGCTTCTCTAATCC	907	
Qy	896	AG	GAGTGGCGAGAT	CAGCTATGGCAGTTCTCT	ACTGAGCTTCTGTCGACAGCTCCAA	955	
Db	908	AG	TAGTGGCCAGAT	CCAGCTGTGGCAGTTCTCT	GTAGAACTCCTGTCTGACAGCTCCAA	967	
Qy	956	CT	CAACTGCATC	ACCTGGGAGGCACAATGGGAGTTCA	AGATGACAGACGACCTGATGA	1015	
Db	968	CT	CAACTGCATC	ACCTGGGAGGCACAACGGGGAGTTCA	AGATGACAGACGACCGGATGA	1027	
Qy	1016	AG	TGGCTCGCGGTT	TGGGGAGAGGAAAGCAAA	CCTAACATGAATATGACAAATCTCAG	1075	
Db	1028	GG	TGGCTCGCGCT	TGGGGAGCGAGAGCAAGCCAA	CATGAATATGACAAATCTCAG	1087	
Qy	1076	CG	TGCATCTTGCT	TACTACTATGCAAAATATTTATGACT	TAAAGTTATGGTAAACGCTA	1135	
Db	1088	CC	GTGCTCGCT	TACTACTATGCAAAAAACATCAT	GCAAGGTGCACGGGAAGCGCTA	1147	
Qy	1136	TG	CTCAAAATTTGAT	TTCCACGGAATCGCTCAGGCGCT	CTCAGCCTCACCTCCAGATC	1195	
Db	1148	TG	CTCAAGTTTGACT	TTCCACGGAATGCGCCAGGCGCT	CTCAGCCTCACCTCCCGAGTC	1207	
Qy	1196	AT	CATGTAACAATAC	CCATCAGACTCCCTCATATGAGT	TCTTACCATGCACACCCCA	1255	
Db	1208	GT	CCCTGTACAAGT	TACCCCTCCGACCTGCGGTAT	CGGTCTCTATCACACCAACCCCA	1267	
Qy	1256	GA	GATGAATTTGT	AGCTCCCCATCCCCCTGCTTT	TGCCGTAACTCATCCAGCTTTT	1315	
Db	1268	GA	GATGAATTTGT	AGCTCCCCATCCCCCTGCGCCCT	CTCCAGTCACTTCTCCAGTTTCTT	1327	
Qy	1316	TG	CTGCCCTTAAT	CCATCTAGTGAATTCACCAACT	TGGAGGATCTTACCCCAATACCAAGCT	1375	
Db	1328	TG	TACCCCGA	CCCATCTAGTGAATTTGCGCAT	TGGGGGATCTTACCCGAACACTAGGCT	1387	
Qy	1376	GCC	AGCTGCTCAT	ATGCTTCCCATCTTGGCACTACTACTA	1417		
Db	1388	CCC	AGCAGCATAT	GCTCTACCTGGGCACCTACTACTA	1429		

RESULT 8	HUMERG2	3166 bp ss-RNA	linear	PRI 08-NOV-1994
LOCUS	Human erg2 gene encoding erg2 protein, complete cds.			
DEFINITION	Human erg2 gene encoding erg2 protein, complete cds.			
ACCESSION	M17254			
VERSION	M17254.1	GI:182186		
KEYWORDS	erg 2 protein.			
SOURCE	erg 2 proteins (human)			
ORGANISM	Homo sapiens			
REFERENCE	1 (bases 1 to 3166)			
AUTHORS	Rao, V.N., Papas, T.S. and Reddy, B.S.			
TITLE	erg, a human ers-related gene on chromosome 21: alternative splicing, polyadenylation, and translation			
JOURNAL	Science 237 (4815), 635-639 (1987)			
MEDLINE	87263429			
PUBMED	3299708			
COMMENT	Original source text: Human cell line COLO 320, cDNA to mRNA, clone lambda 12.			
FEATURES	Location/Qualifiers			
source	1. .3166			
	/organism="Homo sapiens"			
	/mol_type="genomic RNA"			
	/db_xref="taxon:9606"			
gene	/map="21q22.3"			
	257..1645			
	/gene="ERG"			
CDS	257..1645			
	/genes="erg 2 protein"			
	/notes="ERG"			
	/codon_start=1			
	/protein_id="AAA52398.1"			
	/db_xref="GI:182187"			
	/db_xref="GDB:G00-119-884"			
	/translation="MIQTDPDPAHIKEALSVVSDQSLFECAYGTPHLAKTEWASS SDPYGTGSKMSPVQDMLSPQSPVARTIKECNPQVNSRSPDSCSVAKGKMGVGLP SPDTGMGTSYMEEXEHPPTNNTRVIVPADTLMSTDRHVRQMLEWAVKEYGLP DNNVILFQNIIDGKELCKMTKDDPQRLTPSPNADILLSHLYLRETPLHLTSDDDVKA LONSPLMHARNTDLPEVPRSAWTGHGHTPQSKAAQSPSTVPTKTEPQRPOLDPY OILGPTSPRLANDPGSQIOLWQFLLELSDSNSSCITWEGTNGEFPKMTDPDEVARW KERKSPNNYDKLSALAYIYDKNTMKTVHKHRYAKYKDFHGIAGALQHPPESSLY KPSDLPYMGSYFAHFQKMFVAPHFPALFVTSSTFFAANPYNVNSTGGIYFNTRLE TSHMPSHLGTYT"			
repeat_region	291..298			
	/notes="8-bp repeat sequence"			
repeat_region	509..516			
	/note="8-bp repeat sequence"			
polya_signal	3143..3149			
	/notes="putative"			
ORIGIN	227 bp upstream of PvuII site; chromosome 21.			
Query Match	63.6%; Score 921; DB 9; Length 3166;			
Best Local Similarity	81.1%; Pred. No. 1, fe-267;			
Matches 1100; Conservative	0; Mismatches 245; Indels 12; Gaps 2;			
Qy	74 TATTAGGAAGCAATATACAGTGTGTAGTGAAGACCAAGTCCTCTGTTTGAGTGTGCCTACGG 133			
Db	289 TATCAAGGAAGCCCTTATCAGTTGTGAGTGAGGACCAAGTCGTTGTTTGAGTGTGCCTACGG 348			
Qy	134 ATGCCGCCACCTTGCAAGACAGAAATGACAGCTCCTCTTCAGTGAATATGGGCAAC 193			
Db	349 AAGCCNACCTGGCTAAGACAGAGATGACCGGTCTCTCCAGCACATATGGACAGAC 408			
Qy	194 ATCAAGAATAGCGCCGCGCTTCCCAGCAGGACTGGTTATTCACAGCCCCGGCCAGAGT 253			
Db	409 TTCAAGATGAGCCACCGCTCCCTCAGCAGGATTGGCTGTCTCAACCCCCAGCAGGGT 468			
Qy	254 TACATTAAATCGAGTGAACCAACACCAAGGTTAATGGTTCAGGGAATTCACCTGATGA 313			
Db	469 CACCATCAAAATGGAATGTAACTCCCTAGCCAGGTGAATGGCTCAAGGAACCTCTCCTGATGA 528			
Qy	314 CTGCAGCGTGGCAAAAGAGGAGGAAAATGCGTTAGCAGTTCACACAATCTTGGGATGAACACTA 373			

[illegible]

Db 1609 CAGCATTATGCGCTTCTCATCTCGGCACTTACTACTAA 1645

RESULT 9  
XLAJ4126

LOCUS XLAJ4126 2012 bp mRNA linear VRT 21-JAN-2000

DEFINITION Xenopus laevis erg gene (erg\_E).

ACCESSION AJ224126

VERSION AJ224126.1 GI:5420047

KEYWORDS ERG gene; transcription factor.

SOURCE xenopus laevis (African clawed frog)

ORGANISM xenopus laevis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; xenopodinae; Xenopus.

1

REFERENCE 1  
Baltzinger, M., Mager-Heckel, A.M. and Remy, P.  
Xl erg: expression pattern and overexpression during development  
plaid for a role in endothelial cell differentiation  
Dev. Dyn. 216 (4-5), 420-433 (1999)

JOURNAL 20095678

MEDLINE 10633861

REFERENCE 2 (bases 1 to 1012)  
Baltzinger, M.  
Direct Submission  
Submitted (24-FEB-1998) Baltzinger M., UPR9005, MWDGD, Centre  
National de la Recherche Scientifique, 15, RUE Rene Descartes,  
67084, FRANCE

FEATURES

source Location/Qualifiers

1..2012

/organism="Xenopus laevis"

/mol\_type="mRNA"

/db\_xref="taxon:8355"

/clone="lambda ZAPII-erg E"

/clone\_lib="lambda ZAPII"

/dev\_stage="stage 30 embryo"

1..2012

/gene="erg"

225..1595

/gene="erg"

/function="Transcription factor"

/codon\_start=1

/protein\_id="CAB46567.1"

/db\_xref="GI:5420048"

/db\_xref="GOA:Q9W629"

/db\_xref="SPTREML:Q9W629"

/translation="MIVQVDPDAHIEKALSVSEDSQSLPCTVGTPLHAKTEMTASS  
SSDYGOTSKMSPRVPOQDWLSQPSRVTIKMGNSPOVNGSRSPDSCIGKSGKMG  
GSDNMGVNGSYIEEKHISPNMTNERRVIVPADPTLWSTDHVRQLEWAIKEYGL  
DVDVLFQNDIGKELCKMTKEDFHLTPSNADILLSHLYLBERGAFIFPNSSVYQ  
DANORIPSRQDLSYEPERRSAWTHNPAPSKASQPSITIVPKTEDPRQLDPIQLGDT  
SSRLANPSQIQIQLWQFLLSDDSSNCSITWEGTNGEFKMDTDFDEVARWGERKNS  
PNMYDKLSRALRYDYKNTIMVKHGRYAKFDFHGIQAALQHPHPSTMYKYPSEL  
PYMSGYHAHPQKNFVAFHPHPALPVTSSSFAPAPNAYWNSPTGSIYENTRLPASHMSES  
HLGYTY"

ORIGIN

Query Match 61.5%; Score 889.6; DB 5; Length 1012;  
Best Local Similarity 79.7%; Pred. No. 5.le-258;  
Matches 1077; Conservative 0; Mismatches 269; Indels 6; Gaps 2;

QY 74 TATTAAGAAGCATTATTCAGTGGTGGTGAAGACAGTCCCTTGTTGAGTGTGCTACGG 133

Db 257 TATCAAAGAAGCTCTTCAGTGGTGGTGAAGATCAGTCGTTATTCAGTGCACCTACGG 316

QY 134 ATCGCCCCACCTTGCAAGAAGACAGAAATGACAGCTCCTCTTCCAGTGAATATGGCAAC 193

Db 317 AACGCCACATCTTACTMAAAGCGAGATGACGCATCTCTCTCCAGTGAATATGGCAAC 376

QY 194 ATCAAAGATGACCGCGCGTTCCTCCAGCAGACTGGTTATCACAGCCCCCGCCAGAGT 253

Db 377 CTCGAAATAGTCCACGTGTGCTCTCAACAAGATGGCTCTCACAGCTCCATCCAGGGT 436



QY 254 TACCATTAAAGATGGAGTGTAAACCAACCAAGGTTAATGGGTCAAGGAATTCACCTGATGA 313  
Db |||||  
QY 437 GACCATCAAGATGGAATGCAACCTAGTCAAGTCAATGGATCAAGAGCTCCCGGATGA 496  
Db |||||  
QY 314 CTGCAGCGTGGCAAAAGAGGAAATGGTTAGCAGTTCAGACATGTTGGGATGAACATA 373  
Db |||||  
QY 497 CTGCAGTATAGGAAGGAAGCAAAATGGGTGGAGGTCGGATAATGTTGGAATGAATTA 556  
Db |||||  
QY 374 TGGAAAGCTACATGGAAGAGAGATATTCGCGCTTCCAAATATGACAAACCAATGAACGAAG 433  
Db |||||  
QY 557 TGGCAGCTACATAGAGAAACACATTTCTCCCAACATGACAAACCAATGAGAGGAG 616  
Db |||||  
QY 434 AGTTATTGTGGCAGCAGATCTAGTTATGGAGCAGACACCATGTAGCGGAGTGGCTGGA 493  
Db |||||  
QY 617 AGTGATTGTACCTGCTCAACCTACTTTGTGGAGCAGACATCATGTGAGCAGTGGCTAGA 676  
Db |||||  
QY 494 GTGGCCAGTGAAGGAGTATGCTTTCAGAGCGTGGACATCTGTGTTGTCAGAAACATGGA 553  
Db |||||  
QY 677 GTGGCCAAATAAAGAAATACGGCTTCAGATGCGATGCTCTCTGTTCCAAACATCGA 736  
Db |||||  
QY 554 TGGAAAGAGTGTGTAAATGACCAAGATGACTTCCAGAGACTCAGCGGAGCTATAA 613  
Db |||||  
QY 737 CGCAAGGAATATGCAAGATGACCAAGAGATTTCCATCGACTCAGCGCAAGCTACAA 796  
Db |||||  
QY 614 CGCAGATATCTCTCTGTCACACTACACTACCTCAGAGAGAGGAGCCACTTTTATTTT 673  
Db |||||  
QY 797 TGTGATATCTCTCTCGCATCTTCACTACTCTCAGAGAAAGAGTGCTCTTTTATTTT 856  
Db |||||  
QY 674 TCCAAATACATCATGTTTACCAGAGCAACGCAAGAAATAACCAAGGCCAGCATTTACC 733  
Db |||||  
QY 857 CCCAAATCTATCCGTTTATCAAGAGCAACCAAGAAATCCCAAGAGGCCAGGCAAGATTATC 916  
Db |||||  
QY 734 TTATGACAGAGCGAGGAGATCAGCGTGAGGAGTGCAGAGCATCCCACTCAGTCAAAAGC 793  
Db |||||  
QY 917 ATATGAGCCCTCAAGAGAGATCTGTTTGCAAAATCA---TCTGACACCATCAAAAGC 973  
Db |||||  
QY 794 TACCAACCATCATCTTCAACAGTGCACCAAGAGAGAGAGAGCGTCTCAGTTAGATCC 853  
Db |||||  
QY 974 ATCTCAAC---CATCTACACAGTTTCCCAAAACAGAGAGACCCAGGCAAGCTAGATCC 1030  
Db |||||  
QY 854 TTATCAGATTCTTGGACGACAGACCGCTTTCGAAATTCAGAGGAGTGGGAGATACA 913  
Db |||||  
QY 1031 TTATCAGATTCTTGGGCAACAGAGCCGACTTGCAAATTCAGGAAAGTGGGAGATCA 1090  
Db |||||  
QY 914 GCTATGCGAGTCTCTCTGAGCTTCTGTCGAGCAGCTCCAACTCCAACTGCATCACCTG 973  
Db |||||  
QY 1091 GCTCTGCAATCTTATGGAATCTGCTTGGATAGTTCACATCCAACTGCATTTACTTG 1150  
Db |||||  
QY 974 GGAGGCAAAATGGGAGTTCAAGATGACAGACCCCTGATGAAGTGGCTCGCGTGGGG 1033  
Db |||||  
QY 1151 GGAAGGAACCAATGGAGAAATTAAGATGACCGATCTCTGATGAAGTGGCAAGAGCTTGGG 1210  
Db |||||  
QY 1034 AGAGAGAAAGCAAACTAACTGAATGACAACTCAGCGTGCATTCGCTACTA 1093  
Db |||||  
QY 1211 GGAGAGAAAGCAAACTCAACATGAATGATGAAGTGCAGCGTGCATTCGTTACTA 1270  
Db |||||  
QY 1094 CTATGACAAAAATATGACTAAAGTTCTATGGTAAACGGTATGCGCTACAAATTTGATTT 1153  
Db |||||  
QY 1271 CTAGATAAAATATATGACTAAAGTCCATGGCAAGCGCTATGCTTTACAAATTTGATTT 1330  
Db |||||  
QY 1154 CCAGGAATCGCTCAGCGCTCCAGCTCACCCTCCAGAAATCATCCATGACAAATACCC 1213  
Db |||||  
QY 1331 TCATGGGATTCCTCAAGCTCTTCAACCTCATCCATGACATCCAGAGATGAATTTGTTG 1390  
Db |||||  
QY 1214 ATCAGACCTCCCTTACATGATGTTCTACCATGACACACCCCGAGAGATGAATTTGTAGC 1273  
Db |||||  
QY 1391 CTCAGAACTCCATATATGASCTCATACCATGACATCCAGAGATGAATTTGTTG 1450  
Db |||||  
QY 1274 TCCCATCCCGCTGTTTGGCCGTAAACCTCATCCAGCTTTTTCCTGCCCTTAATCCATA 1333  
Db |||||  
QY 1451 CCCCATCCCGCAGCTTTACCTGTGACATCGTCAAGTTTCTTTTGCAGCCCTTAATGCATA 1510  
Db |||||  
QY 1334 CTGGAATTCACCAACTGGAGGCACTTACCCCAATACCAGGCTGCCAGCTGCTCATATGCC 1393  
Db |||||

Db 1511 CTGGAATTCACCAACTGGAAGTATTTATCCAAATACTCGGCTGCCAGCTAGCCATATGTC 1570  
QY 1394 TTCCCATCTTGGCAGCTACTACTAAGTGGGA 1425  
Db 1571 TTCTCATCTTGAACCTACTATTAATCAAGA 1602  
RESULT 10  
AB073078 2131 bp mRNA linear ROD 17-OCT-2001  
LOCUS Mus musculus Erg mRNA, mouse homolog of Human ets-related gene ERG,  
DEFINITION transcript variant 1, complete cds.  
ACCESSION AB073078  
VERSION AB073078.1 GI:16191716  
KEYWORDS  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 Ozawa,R., Noguchi,H., Taylor,T.D., Takeda,T., Hattori,M. and Sakaki,Y.  
Mus musculus Erg mRNA  
2 Published Only in Database (2001)  
2 (bases 1 to 2131)  
REFERENCE 2 Ozawa,R., Noguchi,H., Taylor,T.D., Takeda,T., Hattori,M. and Sakaki,Y.  
TITLE Direct Submission  
JOURNAL Submitted (16-OCT-2001) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suihito-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@gs.c.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)  
FEATURES  
Location/Qualifiers  
1..2131  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/chromosome="16"  
/clone="Erg-1"  
1..2131  
/gene="Erg"  
148..1536  
/gene="Erg"  
/note="homolog of human ets-related gene (ERG), transcript variant 1"  
/codon\_start=1  
/protein\_id="BAB69948.1"  
/translation="MIQTVDPFAAHKEALSVDQSLFECAYGTPHLAKTEMTASS SSDYQTSKSPRVFPQDMLSQAPARVTKMECPNSQVNGSRNSPDCSVNKGKQVGV DVDVLLFONIDGKELCKMTKDDFORLTPTSNADILLSHLYLRETPLPLHSTDVDYDKA LQNSPRIMEARNTDLPYEPERRSAGTSHSLTPOSKAQPSFSAVPTKEDORQPDQPY GILGTSPELLANPGSGOILMQFLLELSDSNCSNCTWEGTNGFKMTDPDEAVRRW GERKSPNNVYDKLSRALRYYYDKNIMTKVHKRYAYKDFPHGIAQALQHPPESSLY KPSDLPTNGSYTHAHPQKNFVSPHPALPVTSSFFSPSPYMNSTPGGIYPTNRLP ASHMSHLGTY"  
ORIGIN  
Query Match 60.9%; Score 881.6; DB 10; Length 2131;  
Best Local Similarity 79.3%; Pred. No. 1.4e-255;  
Matches 1075; Conservative 0; Mismatches 269; Indels 12; Gaps 2;  
QY 74 TATTAAGGAAGCATTATCAGTGGTGAAGACCAAGTCTTCTTGTGAGTGCGCTACGG 133  
Db 180 TATTAAGGAGGCTTGTGAGTGTGAGCGAGGAGGAGTCACTATTGAGTGTGCTACGG 239  
QY 134 ATCGCCCACTTGAAGACAGAAATGACAGCTCTCTTCCAGTGAATATGGCAAC 193  
Db 240 AAGCCACACCTGGTGAAGACAGAGATGACCGCATCTCTTCCAGTGACTATGGCCAG 299



QY 194 ATCAAGATGAGCCGCGTCCCGAGCAGGACTGGTTATCAAGCCCGCGGCGAGGT 253  
 Db 300 ATCAAGATGAGTCCCGAGTCCCTCAGCAGGACTGGCTGTCTCAAGCCCGCGGCGAGGT 359  
 QY 254 TACCATTAAAGATGAGTGAACCAACCAAGTTAAATGGGTCAAGGAATTCACCTGATGA 313  
 Db 360 CACCATCAAGATGAGTGAACCAACCAAGTTAAATGGGTCAAGGAATTCACCTGATGA 419  
 QY 314 CTGACGCTGCAAAAGAGGAGGAAATAGTTAGCAGTTTCAGACAAATGTTGGATGAACCTA 373  
 Db 420 GTGAGTGTGAACAAAGGTGGAGAGATGGTGGGAGCCCGGATCTGTGGGATGAGCTA 479  
 QY 374 TGGAGCTACATGGAAGAGAGCAATATCCCGCTCCAAATATGACCAACCAATGAACGAAG 433  
 Db 480 CGCAGCTACATGGAAGAGAGCAATATCCCGCTCCAAATATGACCAACCAATGAGCGAG 539  
 QY 434 AGTTATTGTCGACGAGATCTAGCTTATGAGACACAGACCATGTACGGCAGTGGCTGGA 493  
 Db 540 AGTGATCTCTCTGACAGATCTCTCTGTGAGACACAGACCATGTCCGACAGTGGCTGGA 599  
 QY 494 GTGGCAGTGAAGAGTATGTCTTCCAGACGTGGACATCTTGTGTTCCAGAACATGA 553  
 Db 600 GTGGCGGTGAAGATATGAGCTTCCGATGTGGAGCTTACTATTTTCAAGATATCGA 659  
 QY 554 TGGGAAAGAGTTGTGTAATAATGACAAAGATGATTCAGAGATCTACGGCGAGCTATAA 613  
 Db 660 TGGGAAAGAGTTGTGTAATAATGACAAAGATGATTCAGAGATCTACGGCGAGCTATAA 719  
 QY 614 CGCAGATCTCTCTGTCACACCTACACTCTCAGAGAGA-----GAGGAGCCAC 664  
 Db 720 TGCCGACATCTCTCTCAGATCTCTCAGAGAGACTCCCTTCCACATCTGAC 779  
 QY 720 TTTTATTTTCCAAATACATCAGTTTACCCAGAGCAACGCAAGAAATAACAAGGCC 724  
 Db 780 TTCCGATGACGTTGATAGGTTTACAAACTCTCCAGGTTAATGATGATCCAGAAACAC 839  
 QY 725 AGATTACTTATGAGCAAGGAGAGAGATGAGCTGAGAGAGTCAAGAGCATC---CCAC 781  
 Db 840 AGATTACTTATGAGCTTCCAGAGATGAGCTGAGAGAGTCAAGAGCATCAGCTTCACTCC 899  
 QY 782 TCAGTCAAAAGTACCCCAACCATCATCTTCAACAGTGCCTCAAAAGAGAGAGAGCGTCC 841  
 Db 900 TCAGTCAAAAGTCTGACGATCTCTCTGAGTGCCTCAAAAGAGAGAGAGCGTCC 959  
 QY 842 TCAGTTAGATCTTATGAGTTCTTGGACCAAGAGAGAGCGTCTTGAATTCAGAGAG 901  
 Db 960 TCAGTTAGATCTTATGAGTTCTTGGACCAAGAGAGAGCGTCTTGAATTCAGAGAG 1019  
 QY 902 TGGGAGATACAGCTATGAGCTTCTTACTGAGTCTTCTGAGAGAGTCTTCAACTCCAA 961  
 Db 1020 TGGCAGATCCAGCTGTGAGTCTTCTGAGTCTTCTGAGAGAGTCTTCAACTCCAA 1079  
 QY 962 CTGCATACCTGGGAGGCAAAATGGGAGTTCAGATGACAGACCTGATGAGTGC 1021  
 Db 1080 CTGCATACCTGGGAGGCAAAATGGGAGTTCAGATGACAGACCTGATGAGTGC 1139  
 QY 1022 TCGCGGTTGGGAGAGAGAAAGCAACCTAAATGAGTCTTCAACTATGAGAGTCAAGCTGC 1081  
 Db 1140 TCGCGGTTGGGAG 1199  
 QY 1082 ACTTCGCTACTACTATGACAAATATATGATGATGATGATGATGATGATGATGATGATGAT 1141  
 Db 1200 CTTCCGCTACTACTACGACAAATATATGATGATGATGATGATGATGATGATGATGATGAT 1259  
 QY 1142 CAAATTTGATTTCCAGGAGTCTGAGGCTCCAGCTCCAGCTCCAGCTCCAGCTCCAGCT 1201  
 Db 1260 CAAATTTGATTTCCAGGAGTCTGAGGCTCCAGCTCCAGCTCCAGCTCCAGCTCCAGCT 1319  
 QY 1202 GTACAAATACCCATCAGACCTCCCTTACATGATGATGATGATGATGATGATGATGATGAT 1261  
 Db 1320 GTACAAATACCCCTCCGAGCTGACATATGAGGCTTCTTATCAGGCTCCATATCAGGCT 1379  
 QY 1262 GAACTTTGATGCTCCGATCCCTTCTGAGGCTTCTGAGGCTTCTGAGGCTTCTGAGGCT 1321

Db 1380 GAACCTTTGTTGTTCTCCCAACCTTCCCGCTCCCACTCCCACTCCCACTTTTTCCTC 1439  
 QY 1322 CCTAATCATCTGGAATTCACCAACTGGAGGATCTACCCCAATACAGGCTCCGAGC 1381  
 Db 1440 CCGAACCACATCTGGAATTCACCAACTGGAGGATCTACCCCAATACAGGCTCCGAGC 1499  
 QY 1382 TGCTCATATGCTTCCCATCTTGGCAGCTTACTACTA 1417  
 Db 1500 CAGCCATATGCTTCTCCTTCCCTGGGACCTTACTACTA 1535

RESULT 11  
 AB073080 2209 bp mRNA linear ROD 17-OCT-2001  
 LOCUS Mus musculus Brg mRNA, mouse homolog of Human ets-related gene BRG,  
 DEFINITION transcript variant 3, complete cds.  
 ACCESSION AB073080  
 VERSION AB073080.1 GI:16197544  
 KEYWORDS

SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 Ozawa,R., Noguchi,H., Taylor,T.D., Takeda,T., Hattori,M. and Sakaki,Y.  
 TITLE Mus musculus Brg mRNA  
 JOURNAL Published Only in Database (2001)  
 REFERENCE 2 (bases 1 to 2209)  
 AUTHORS Ozawa,R., Noguchi,H., Taylor,T.D., Takeda,T., Hattori,M. and Sakaki,Y.

TITLE Direct Submission  
 JOURNAL Submitted (16-OCT-2001) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), 1-7-22 Suehiro-chou, Tsukuba, Ibaraki, Japan, 305-8565, Japan (E-mail: hattori@gs.c.riken.go.jp, URL: http://hgp.gs.c.riken.go.jp/, Tel: 81-45-503-9111, Fax: 81-45-503-9170)  
 FEATURES Location/Qualifiers

source 1..2209  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /chromosome="16"  
 /clone="Brg-3"  
 1..2209  
 /gene="Brg"  
 148..1608  
 /note="homolog of human ets-related gene (ERG), transcript variant 3"  
 /codon\_start=1  
 /protein\_id="BAB69950.1"  
 /db\_xref="GI:16197545"  
 /translation="MIQTPDPAHIIKEALSIVSEDSQSLFBCAYGTPHLAKTEMTASS  
 SSQYQTKMSRVPQDQWLPQAPARVITKMECPNSQVNGSRNSDECSVNGKRVG  
 SPDYVMSYSGYMEKHYVPFNNMTNERRIVPADPTLSTHDVQWLEWAVEYDGLL  
 DVIDLLFNIDGKELCKMTKDFORLTSPYNADILLHLVRLTFLPHLTSDDDVKA  
 LONSRLHARTGGAATFFNTSVVPEATORITTRPDLPEPPRPSAWTGHSLTPQ  
 SKAAPSAPKPTEDQRPOLDPQIOLGTSRLANPGSGIOLWQFLLELLSDSSNS  
 NCITWEGNGFKMTDPEVARWGERKSKPMNYDKLSRLALRYVDKNITKVIKGR  
 YAKDFDFGIAQALQHPPESSLYKIPSDLPYMGSHAHQKNNFVSPHPALPTSS  
 SFFASPNFYWNSPTGGIYENTRLPASHMPSHLGTY"

gene  
 CDS  
 ORIGIN  
 Query Match 59.9%; Score 867.2; DB 10; Length 2209;  
 Best Local Similarity 77.8%; Pred. No. 3.2e-251;  
 Matches 1111; Conservative 0; Mismatches 233; Indels 84; Gaps 2;  
 QY 74 TATTAAGGAGCATTTATCAGTGTGAGTGAAGACAGGCTCTTGTGAGTGTGCTACGG 133  
 Db 180 TATTAAGGAGGCTTGTCTAGTGTGAGGAGGACCACTTATTTGAGTGTGCTACGG 239

134	QY	ATCGCCCCACTTGTGCAAGA	CAGAAATGA	CAGCGCTCTCTTCCAGTGAATATGGGCAAC	193	
240	Db	AACGCCACACCTGGCTAAGA	CAGAGATGACGGCATCTCTTCCAGTACTATGGCCAGAC	299		
194	QY	ATCAAGAATGAGCCCGCGCTT	CCCCAGCAGGACTGGTTAT	CAAGCCCGGCCAGAGT	253	
300	Db	ATCCAAGATGAGTCCAGAGAT	CCCTCAGCAGGACTGGCTGTCT	CAAGCCCCCAGCCAGGGT	359	
254	QY	TACCATTAAAGATCGAGTGT	TAACCCCAAAC	CCAGGTTAATGGGTCAAGGAATTCACCTGATGA	313	
360	Db	CACCATCAAGATGGAGTGAAC	CCTAGTCAAGTGAATGGTTC	CAGGAACATCACTGATGA	419	
314	QY	CTGACCGGTGGCAAAAGGAG	GGAAATGGTTAGCAGTTC	CAGACAAATGTTGGGATGAACATA	373	
420	Db	GTCCAGTGTGAACAAAGGT	GGGAAGATGGTGGGCAGCCCGGAT	CTACTGTGGGATGAGCTA	479	
374	QY	TGGAAGCTCATATGGAAGAG	AAGCATATTCGCGCTTCAAA	TATGACACCAACATGAACGAAG	433	
480	Db	CGCAGCTCATGAGGAGAG	CAATGTCGCGCTCCCAAT	TATGACCACAATGAGCGCAG	539	
434	QY	AGTTATTGCCAGCAGATCCT	ATAGTTATGGAGCAGACAC	ATGTCAGGCTGGCTGGGA	493	
540	Db	AGTGATCGTCCCTGCGAGAT	CCTACTCTGTGGAGCAGAC	CCATGTCGACAGTGGCTGGA	599	
494	QY	GTGGGAGCTAAAGAGTATGG	TCTTCCAGACGTGGACATCT	GTGTGTTCCAGAACATGGA	553	
600	Db	GTGGCGGTGAAGAATA	TGGCTCTCTGATGTGGAC	GTCTTACTATTCAGAAATATCGA	659	
554	QY	TGGGAAGAGTTGTGTAAAT	TGACCAAGATGACTTCCAGAGAT	CTCACGCCGAGCTATAA	613	
660	Db	TGGGAAGGAGCTGTGCAAG	ATGACAAAGGATGACTTCCAGCGG	CTCACGCCGAGCTACAA	719	
614	QY	CGCAGATATCTCTGTGTG	CACACCTACACTACCTTCAGAGAG	-----	654	
720	Db	TGCGGACATTCTTCTCTCA	TCTCCACTACCTCAGAGAGACT	CCCTCCCATCTGAC	779	
655	QY	-----	-----	-----	654	
780	Db	TTCCGATGAGTTGATAAGG	CTTTACAAAAC	TCTCCACGGTTAATGATGCCAGAAACAC	839	
655	QY	--GAGGAGCACATTTATTT	TTTCCAAATACATCAGTTT	TACCGAGACCAACGAAGAAT	712	
840	Db	AGGGGTGACGCTTTTATTT	TTCCAAATATCTTCAGTATAT	CCCGAAGCTACGCAAAAGAAT	899	
713	QY	AACAACAAGCCAGATTTAC	CTTTATGAGCAAGCGGAGAGAT	CAGCGTGGACGAGTCAACAG	772	
900	Db	TACAAC	TAGSCCAGATTTAC	CTTTATGAGCTCCCGAGAGATCAGCCTGGACCGGCCACAG	959	
773	QY	CCATC---	CCACTCAGTCAAAAGCTAC	CCAAACCATCATCTTCAACAGTGGCCAAAACAGA	829	
960	Db	CCACCTCACCCCTCAGT	CCAAAGCTGCTCAGCCATCT	CCCTCTGCACTGTGCCAAAACATGA	1019	
830	QY	AGAC	CAGGCTCCTCAGTTAGAT	CCCTTATCAGATTCTTGGACCGAC	CAGCAGCGCTTTGC	889
1020	Db	AGAC	CAGGCTCCTCAGTTAGAT	CCTTACAGATCTCGGACCGAC	CAGTACCGCCTTGC	1079
890	QY	AAAT	TCAGGAGGTGGGAGATAC	CAGCTATGGCAGTTCTTACTGGAGCTCTGTGGACAG	949	
1080	Db	TAAT	TCAGGTAGTGGCCAGAT	CCAGCTGTGGCAGTTCCCTGCTCGAATCCTCTGT	CAGACAG	1139
950	QY	CTCCAACT	CCAACTGATCACTT	CGGAGGCAAAATGGGGAGTTC	CAAGATGACGACCC	1009
1140	Db	CTCCAACT	CCAACTGATCACTT	CGGAGGCAAAATGGGGAGTTC	CAAGATGACGACCC	1199
1010	QY	TGAT	GAAGTGGCTCGGCGTT	GGGGAGAGGAAAAGCAAACT	TAAATGAACTATGACAA	1069
1200	Db	GGAC	GAGTGGCTCGGCGCT	GGGGAGAGGAAAGACCAACCC	CAACATGAATATGACAA	1259
1070	QY	ACT	CAGCGGTGCACTTCGCTACT	ATATGACAAAATATTTATGACT	TAAAGTTCAATGGTAA	1129
1260	Db	GCT	CAGCGCGCCCTCGCTACT	ACTACGACAAAACATCATGAC	CAAGGTGACGGAA	1319
1130	QY	ACG	CTATGCTCAAAATTTGAT	TTTCAACGGAATCGCTCAGGCGCT	CTCAGGCTCAACCTCC	1189

Db	1320	GC	GCTACGCCCTACAAGTTTGAC	TTCCACGGGATTGCCAGGCCCTGCAGGCCACCCCTCC	1379
Qy	1190	AG	ATCATCATCATGTACAAAT	TACCATCAGACCTCCCTCATCATGATGTTCTCTACCATGCACA	1249
Db	1380	TG	AGTCGTCCTGTACAAG	TACCCCTCCGACCTGGCCATACATGGGTCTCTATCAGCCCA	1439
Qy	1250	CC	CCAGAAGATGAAC	TTTGTAGCTCCCATCCCCCTGCTTTTCCCGTAACTCATCCAG	1309
Db	1440	CC	CCAGAAGATGAAC	TTTGTGTCTCCCACTCCCGTCTCCAGTCACATCTTCCAG	1499
Qy	1310	CT	TTTTTGCTGCCCTTAAT	CCATACCTAGTGGATTCCACCACTGGAGGCATCTACCCCAATAC	1369
Db	1500	TT	CTTTTGCTTCCCGAA	CCCATCTAGGATTCCACGACTGGGGGATCTACCCGAAAC	1559
Qy	1370	CAG	CTGCCAGCTGCTCAT	ATGCTTCCCATCTTGGCACCTACTACTA	1417
Db	1560	TAG	CTCCAGCAGCATAT	GCCCTCTACCTGGGCACCTACTACTA	1607
RESULT 12					
AX657230					
LOCUS	AX657230		1291 bp	DNA	linear
DEFINITION	Sequence 3 from Patent WO03000724.				PAT 22-MAR-2003
ACCESSION	AX657230				
VERSION	AX657230.1		GI:29160008		
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
FEATURES					
source					
CDS					
ORIGIN					
Query Match					
Best Local Similarity					
Matches					
Qy	294	TCA	GGAATTCACCTGATG	ACTCCAGCGTGGCAAAAGGAGGAAATGGTTACAGTTCA	353
Db	101	TCA	GGAATTCCTCTGAT	GAAATCAGTGTGGCCAAAGGGGGAAGATGGTGGCGACGCCA	160
Qy	354	GAC	AAATGGGGATGAC	TATGGAAGTATGGAAGAGGAGCATATTCGGCTCCCAAT	413
Db	161	GAC	AAATGGGGATGAC	TATGGAAGTATGGAAGAGGAGCATATTCGGCTCCCAAT	220
Qy	414	AT	GACAAACCAATGA	ACGAAGATTTTGTGCCAGCAGATCCCTACGTTATGGACACAGAC	473
Db	221	AT	GACCAACGAACG	AGCGCAGAGTTTCGTGCCAGCAGATCCCTACGTTATGGACACAGAC	280

Qy	474	CATGTACGGCAGTGGCTGGAGTGGCGAGTGNAGAGATATGGTCTTCAGACGTGGACATC	533
Db	281	CATGTGGCGAGTGGCTGGAGTGGCGGTGAAGAATATGGCTTCAGACGTCAACATC	340
Qy	534	TTGTGTTTCAGAAACATTGTATGGGAAGAGTGTGTAAATATGACCAAGATGACTCCAG	593
Db	341	TTGTATTCAGAAACATCGATGGGAAGAACTGTGCAAGATGACCAAGGACGACTTCAG	400
Qy	594	AGACTCAGCGGAGCTATTAACGAGATATCTCTCTGTTCACACTTACTACTCTCAGAGAG	653
Db	401	AGGCTACCCCGGAGCTCAACGCGGACATCTCTCTCACAATCTCCACTACTCTCAGAGAG	460
Qy	654	AGAGGAGCCACTTTTATTTTTCAAAATACATCAGCTTTACCCAGAAAGCAACGCCAAAGAATA	713
Db	461	AGGGGTGAGCTTTTATTTTCCCAATACCTTCAGTATATCTGAAGTACGCAAGAATTT	520
Qy	714	ACAAACAAGGCGAGATTTACCTTTATGAGACGAGGAGAGATCAGCGTGGAGAGTCAAGC	773
Db	521	ACAACTAGGCGAGATTTACCATATGAGCCCCCAGAGAGATCAGCCTGGAGCGGTACGGC	580
Qy	774	CATCCCA--CTCAGTCAAAAGCTACCCCAACCATCTTCAACAGTGGCCCAAAACAGAA	830
Db	581	CACCCACGCCCCAGTGAAGCTGCTCAACCATCTCTTCCACAGTGCCTCAAAACTGAA	640
Qy	831	GACCAGCGTCTCAGTTAGATTCCTTATCAGATTTCTGGACCGACCCAGCAGCGCTTGCA	890
Db	641	GACCAGCGTCTCAGTTAGATTCCTTATCAGATTTCTGGACCCAACTAGTAGCGCTTGCA	700
Qy	891	AATCCAGGAGTGGCGAGATACAGCTATGGCAGTTCCTACTGAGAGTCTCTGTGGAGCAGC	950
Db	701	AATCCAGGAGTGGCGAGATCCAGCTTTGGCAGTTCCTCTGGAGTCTCTGTGGAGCAGC	760
Qy	951	TCCAACTCCAACTGCATCAGCTGAGGAGGCGACAAATGGGGAGTTCAAGATGACAGACCCCT	1010
Db	761	TCCAACTCCAGCTGCATCAGCTGAGGAGGCGACCAACGGGGAGTTCAAGATGACAGATCCC	820
Qy	1011	GATGAAGTGGCTCGGGCTTGGGAGAGAGGAAAGCAAACTTAATGACTTAAAGTTTCATGGTAAA	1130
Db	821	GACGAGGTGGCCGCGCTGGGAGAGCGGAAGAGCAAACTTAATGACTTAAAGTTTCATGGTAAA	880
Qy	1071	CTCAGCGGTGCACATTCGTACTACTATGACAAAAATATTATGACTTAAAGTTTCATGGTAAA	1130
Db	881	CTCAGCGCGCGCTTCGTTACTATATGACAAAGAACTATGACCAAGGTCCATGGGAG	940
Qy	1131	CGCTATGCCTACAAATTTGATTTCCACGGAAATCGCTCAGGCGCTTCAGCCTCAACCTCCA	1190
Db	941	CGCTACGCTACAAAGTTCGACTTCCACGGGATCGCCAGGCGCTTCAGCCTCAACCTCCA	1000
Qy	1191	GAATCATCCATGTACAAATACCCATCAGACCTCCCTTACATGAGTTCCTACCATGCACAC	1250
Db	1001	GAGTCATCTCTGTACAAAGTACCCCTCAGACCTCCCGTACATGGGCTCTCTATCAGCCAC	1060
Qy	1251	CCCCAGAGAATGAATTTGTAGCTCCCCATCCCTCTGTTTGGCCGCTAACTCATCCAGC	1310
Db	1061	CCACAGAAGATGAATTTGTGGCGCCCCACCTCCAGCGCTCCCGGTGACATCTTCCAGT	1120
Qy	1311	TTTTTTGTGTCCTTAATCCATCTGGAATTTACCAACTGAGGAGATACCCCAATACC	1370
Db	1121	TTTTTTGTGTCCTCAAAACCCATCTGGAATTTACCAACTGAGGAGATACCCCAATACC	1180
Qy	1371	AGGCTGCCAGTGTCTCATATGCCCTTCCCATCTTGGCACTACTACTAA	1418
Db	1181	AGGCTCCCAACAGCCATATGCCCTTCTCATCTGGGCACTTACTACTAA	1228

RESULT 13

AB073079

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

AB073079

Mus musculus Erg mRNA, mouse homolog of Human ets-related gene ERG, transcript variant 2, complete cds.

AB073079.1

GI:16197542

2133 bp

mRNA

linear

ROD 17-OCT-2001

SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1  
 AUTHORS Ozawa,R., Noguchi,H., Taylor,T.D., Takeda,T., Hattori,M. and Sakaki,Y.  
 TITLE Mus musculus Erg mRNA  
 JOURNAL Published Only in Database (2001)  
 REFERENCE 2 (bases 1 to 2133)  
 AUTHORS Ozawa,R., Noguchi,H., Taylor,T.D., Takeda,T., Hattori,M. and Sakaki,Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-OCT-2001) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@gsc.riken.go.jp URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, fax:81-45-503-9170)  
 FEATURES  
 source  
 1..2133  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /chromosome="16"  
 /clone="Erg-2"  
 1..2133  
 /gene="Erg"  
 149..1540  
 /gene="Erg"  
 /notes="homolog of human ets-related gene (ERG), transcript variant 2"  
 /codon\_start=1  
 /protein\_id="BAB69949.1"  
 /db\_xref="GI:16197543"  
 /translations="MIQTDPDPAHIKEALSVVSEDSQSFCEAYGTFPLAKTEWTAS  
 SSYQGTSKMSPVQDDMLSQPARVITKMECPNSQVNSRNSPDECSVNGKGMVGV  
 SPDTGMSYGSWEERKHVPNNVTNRRVIVPADTLWSTDHVRWLWNAVEYGLL  
 DVNVLQNIIDGKELCKMFKDDPRLTPSNADILLHLHLRETPLPHLTSDDVKA  
 LQNSPLMHRANITGGAAFPFPNTSVPEATORITTPPAQSPSPAYPEKTDQRPOLDP  
 YQLIGTSSKLANPGSGQQLQWFLLELSDSSNSNCITWEEINGEFKMTDPDEVARL  
 WGRSKLPNNYDQKLGRALRYDYDKNIMTKVHGKRAYKDFDFHGAQALQPHHPPESSL  
 YKYPISDPMYGSYHAHPQKNFYSPHPPALPVTSSFFASPNFYNWSPGTGIIYPTNLEL  
 PASHMPSHLGTYT"  
 ORIGIN  
 Query Match 57.2%; Score 828; DB 10; Length 2133;  
 Best Local Similarity 76.8%; Pred. No. 2.5e-239;  
 Matches 1046; Conservative 0; Mismatches 295; Indels 21; Gaps 2;  
 QY 74 TATTAAAGGAAGCATATCAGTGGTGAAGTGAAGACCAAGTCCCTGTGTTGAGTGTGCTACGG 133  
 DB 181 TATTAAAGAGCCTTGTGAGTGTGAGCGAGGACCAAGTCACTATTGAGTGTGCTACGG 240  
 QY 134 ATGCCCCACCTTGTCAAAGACAGAAATGACAGCCTCTCTTCAGTGAATATGGGCAAC 193  
 DB 241 AACGCCACACTGGCTAAGACAGAGATGACCGCATCCTCTTCCAGTCACTATGGCCAGAC 300  
 QY 194 ATCAAGATGAGCCCGCGGTTCGCCAGCAGGACTGGTTATACAGCCCCCGGCCAGAGT 253  
 DB 301 ATCCAAGATGATCCAGAGTCCCTCAGCAGGACTGGCTGTCTCAAGCCCCCAGCCAGGT 360  
 QY 254 TACATTAAAGATGAGGTGTAAACCAAAACCAAGTTAATGGGTCAAGGAATTCACCTGATGA 313  
 DB 361 CACCATCAAGATGAGGTGCAACCCCTAGCTAGGTGAATGGTTCCAGGAACTCACCTGATGA 420  
 QY 314 CTGCAGCGTGCRAAAGAGGGAAAAATGGTTAGCAGTTTCAGACAATGTGGGATGAACCTA 373  
 DB 421 GTGCAGTGTGAAACAAAGGTGGGAAGATGTTGGGACGCCCGGATACTGTGGGATGAGCTA 480  
 QY 374 TGGAAAGCTTACATGGAAGAGAAGCATATTCGCCCTCCAAATATGACACCAATGAACCAAG 433  
 DB 481 CGGCAGCTTACATGAGCAGGAAGCATGTGCCCTCCCAATATGACCAACAATGAGCGCAG 540

## ORIGIN

Query Match 57.2%; Score 828; DB 10; Length 2133;  
Best Local Similarity 76.8%; Pred. No. 2.5e-239;  
Matches 1046: Conservative 0; Mismatches 295; Indels 21

Qy	74	TATTAAAGGAAGCATATACAGTGTGTGAGTGAAGACCAAGTCTCTTGTTTGAGTGTGCCTACGG	133
Db	181	TATTAAAGGAGCCCTGTGCAGTGTGAGCGAGGACCAAGTCACTATTGTGAGTGTGCCTACGG	240
Qy	134	ATFCGCCCCACCCTTCGAAAGACAGAAATGACAGAGCTCTCTTCACAGTGAATATGGGCAAAAC	193
Db	241	AACGCCACACCTGGCTTAAGACAGAGATGACCGCATCTCTTCAGTGAATATGGCCAGAC	300
Qy	194	ATCAAGAAGATAGCCGCGGCTTCCCAGCAGGAGCTGGTTATCACAGCCCCCGGCCAGAGT	253
Db	301	ATCCGAAGATGAGTCCCAGAGTCCCTCAGCAGGAGCTGCTGTCTCAAGCCCCCAGCCAGGCT	360
Qy	254	TACCATTAAGATGGAGTGTAACCCCAAACCCAGGTAAATGGGTCAAGGAAATTCACCTGATCA	313
Db	361	CACCATCAAGATGGAGTGCAACCCCTAGTCAGGTGAATGGTTCCAGGAACTCACCTGATGA	420
Qy	314	CTGCAGCTGCGCAAAAGGAGGAAAATGGTTACAGTTTCAGAGCAATGTTTGGGATGAATCA	373
Db	421	GTGCAGTGTGAACAAAGGTGGGAAGATGGTGGGAGCCCGGATCTGTGGGGTGAAGCTTA	480
Qy	374	TGGAAGCTACATGGAAGAGAGCATATATCCGCTCCCAATATATGACAAACCAATGAACGAAG	433
Db	481	CGCGAGCTACATGAGAGAGAGCATGTGCGGCTCCCAATATGACCAACAAATGAGCGCAG	540

RESULT 13	AB073079	2133 bp	mRNA	linear	ROD 17-OCT-2001
LOCUS	AB073079				
DEFINITION	Mus musculus Erg mRNA, mouse homolog of Human ets-related Gene ERG, transcript variant 2, complete cds.				
ACCESSION	AB073079				
VERSION	AB073079.1				GI:16197542
KEYWORDS					

QY 434 AGTTATTGTCAGAGATCTTACGTTATGGAGACAGACCATGTAGCGAGTGGCTGGA 493  
 Db |||||  
 QY 541 AGTGATGCTCCCTGAGATCTTACTCTGTGGAGACAGACCATGTCCGACGAGTGGCTGGA 600  
 Db |||||  
 QY 494 GTGGCAGTGAAGAGTATGCTTCTCAGAGCTGAGACATCTTGTGTTTCCGAACATGGA 553  
 Db |||||  
 Db 601 GTGGCGGTGAAGAATATGGCTCTCCGATGTGAGCTCTTACTATTTCAGAAATATCGGA 660  
 QY 554 TGGGAAAGAGTGTGTAAATGACCAAGATGACTTCCAGAGACTCAGCGGAGCTATAA 613  
 Db |||||  
 Db 661 TGGGAAGAGGTGTGCAAGATGACAAGATGACTTCCAGGGCTCAGCGGAGCTACAA 720  
 QY 614 GCGAGATATCTCTGTCAGACCTACACTACTCTGAGAGAGAGAGGACCTTTTATTTT 673  
 Db |||||  
 Db 721 TGGCGACATCTTCTCTCACATCTCCACTACCTCAGAGAGACTCCCTTCCACATCTGAC 780  
 QY 674 TCCAAATACATCAGTTTACCAGAGCAA-----CGCAAGAATAAC 715  
 Db |||||  
 Db 791 TTCGATGACGTTGATAGGCTTTACAAACTCTCCAGGTTAATGATGCCAGAACAC 840  
 QY 716 AACAAAGCCAGATTACCTTATGAGCAAGCGAGGAGATCAGCGTGGACGATCAGACCA 775  
 Db |||||  
 Db 841 AGGGGTGAGCTTTTATTTTCCCAATACTTCAGTATATCCGAAGCTAGCCAAAGAT 900  
 QY 776 TCCCACTCAGTCAAAAGCTACCAACCATCTTTCACAGTGCCTCAAAACAGAGACCA 835  
 Db |||||  
 Db 901 TACAACTAGGCC---AGTGTCTAGCCATCTCCCTCTGCACTGCCCAAACTGAAGACCA 957  
 QY 836 GCGTCTCAGTTAGATCTTATCAGATCTTGGACGACAGCAGCGCTTTCGAAATCC 895  
 Db |||||  
 Db 958 GCGTCTCAGTTAGATCTTACCAGATCTGGGACGACAGCAGTAGTGGCGCTTGTATATCC 1017  
 QY 896 AGGAGTGGGAGATACAGCTATGCGAGTTCCTACTGGAGCTTCTGCGGACAGCTCAA 955  
 Db |||||  
 Db 1018 AGGTAGTGGCAGATCCAGCTGTGGCAGTCTCTGCTGGAATCTCTGTGACAGCTCCAA 1077  
 QY 956 CTCCAACTGCATCACTCGGAGGACCAAAATGGGAGTTCAAGATGACAGACCTCGATGA 1015  
 Db |||||  
 Db 1078 CTCCAACTGCATCACTCGGAGGACCAAGCGGAGTTCAAGATGACAGACCTCGAGCA 1137  
 QY 1016 AGTGCTCGGCTGGGAGAGAGAAAGCAAACTTAACATGAATATGACAACTCAG 1075  
 Db |||||  
 Db 1138 GGTGCTCGGCTGGGAGAGAGAGCAAGCAAGCCAAATGAATATGACAACTCAG 1197  
 QY 1076 CCGTGCACTTCGCTACTACTATGACAAAAATATATGACTAAAGTTTATGTTAAAGCTA 1135  
 Db |||||  
 Db 1198 CCGGCTCTCGCTACTACTACGACAAAAACATCATGACCAAGGTGACGGAAGCGCTA 1257  
 QY 1136 TGCCTACAAATTTGATTTTCCAGGAATGCTCAGGCTCTCCAGCTCACCCTCCAGATC 1195  
 Db |||||  
 Db 1258 CGCTTACAAGTTTGACTTCCAGGGATTGCGCAGGCTGCGAGCCCTCAGCTCTGAGTC 1317  
 QY 1196 ATCCATGTACAAATACCATCAGACCTCCCTACATGATGATGTTCTTACCATGACACCCCA 1255  
 Db |||||  
 Db 1318 GTCCCTGTACAAATACCATCAGACCTCCGACCTGCCATACATGGGCTCTTATCAGCGCCACCCCA 1377  
 QY 1256 GAAATGAACCTTTGAGTCTCCCACTCCCTCTGCTTTGCCCCGTAACTCATCAGCTTTT 1315  
 Db |||||  
 Db 1378 GAAGATGAACCTTTGAGTCTCCCACTCCCTCTGCTTTCCAGTCACTCTCCAGTTCTT 1437  
 QY 1316 TGCTGCCCTTAATCCATCTGGAATTCACCAACTGGAGGATCTACCCCAATACAGGCT 1375  
 Db |||||  
 Db 1438 TGCTTCCCGCAACCAATCTGGAATTCACCAACTGGGGGATCTACCCGAACACTAGGCT 1497  
 QY 1376 GCGAGCTGCTCATATGCTTCCCATCTTGGCACTTACTACTA 1417  
 Db |||||  
 Db 1498 CCCAGCCAGCCATATGCTCTCACCTGGGACCTACTACTA 1539

RESULT 14  
 XLAJ4125  
 LOCUS XLAJ4125 1798 bp mRNA linear VRT 21-JAN-2000  
 DEFINITION Xenopus laevis erg gene (erg\_A).

ACCESSION AJ224125  
 VERSION AJ224125.1 GI:5420045  
 KEYWORDS ERG gene; transcription factor.  
 SOURCE Xenopus laevis (African clawed frog)  
 ORGANISM Xenopus laevis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;  
 Xenopodinae; Xenopus.  
 REFERENCE 1  
 AUTHORS Baltzinger M., Mager-Heckel A.M. and Remy, P.  
 TITLE Xl-erg: expression pattern and overexpression during development  
 JOURNAL Plead for a role in endothelial cell differentiation  
 MEDLINE Dev. Dyn. 216 (4-5), 420-433 (1999)  
 PUBMED 20099678  
 REFERENCE 2 (bases 1 to 1798)  
 AUTHORS Baltzinger M.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-FEB-1998) Baltzinger M., UPR9005, WMDCD, Centre  
 National de la Recherche Scientifique, 15, RUE Rene Descartes,  
 67084, FRANCE  
 FEATURES  
 Location/Qualifiers  
 1..1798  
 /organism="Xenopus laevis"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:8355"  
 /clone="lambda ZAPII-erg A"  
 /clone\_lib="lambda ZAPII"  
 /dev\_stage="stage 30 embryo"  
 1..1798  
 /gene="erg"  
 321..1778  
 /gene="erg"  
 /function="transcription factor"  
 /codon\_start=1  
 /protein\_id="GI:5420046"  
 /db\_xref="GI:5420046"  
 /db\_xref="GOA:Q9W700"  
 /translation="VQHWAFPTVEDYSNEALSVVSEDSLEFCTGTGPHLTKTEMTA  
 SSSDIQTSKMSPRVPOQDWLSQPPSRVTIKMNCNPQVNGSRSPDDCSIGKSKM  
 GGSDVNGVNGYSIEERHISPPNNTNRRVIVPADPLWSTIDHVRQWLEWALKEYG  
 LPDVLVLFQNDIGKELCKMTEDFRLTFSYNADILLHLHLYRETPLPHLTSDDDVD  
 KALQSPRLMHARTGGASFIPNSVYQDANQRIPSRQDLSYFSPRSAMTNHPAPP  
 SKAOSPSTVPTKTEDPRDLPDPTGILGPTSSRLANPGSQIQLWQFLLELLSDSSNSN  
 CTTWGTEGFEKMTDPDEVAREWGRKSKPNMVDKLSRALRYVDKIMTKVHGRRY  
 AKFPHGTAQALQHPHPPESTNYKPSBELPYMSSYHAHPQKNFVAPHPALPVTSSS  
 FFAAPNAYNVSPTGISIYENTRLPASHMSHLGTYI"  
 ORIGIN  
 Query Match 55.4%; Score 801.8; DB 5; Length 1798;  
 Best Local Similarity 75.3%; Pred. No. 2.2e-231;  
 Matches 1079; Conservative 0; Mismatches 267; Indels 87; Gaps 3;  
 QY 74 TATTAGGAAGCATATCAGTGTGAGTGAAGACCAAGTCTCTTCCAGTGAATATGGGCAAAAC 133  
 Db 359 TAGTAATGAAGTCTTTTCAGTGTGAGTGAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 418  
 QY 134 ATCGCCCCACCTTGCACCAAGACAGAAATGACAGCTCTCTTCCAGTGAATATGGGCAAAAC 193  
 Db 419 AAGCCACATCTTACTTAAACCGAGATGACCGCATCTTCTCCAGTGAATATGGGCAAAAC 478  
 QY 194 ATCAAGATGAGCCCGCGGTTCCTCCAGCAGAGTCTGTTATCAGACCCCGCCGAGAGT 253  
 Db 479 CTCGAAATGAGTCCACGTGTGCTCAACAAGACTGGCTCTCACAGCTCCATCCAGGT 538  
 QY 254 TACCATTAAGATGAGTGAATCCCAACCGAGTTAATGGTCAAGGAATTCACCTGATGA 313  
 Db 539 GACCATCAAGATGAATGCAACCTAGTCAAGTCAATGATCAAGAGGCTCCCGGATGA 598  
 QY 314 CTGACGCTGGCAAAAGGAGGAAATGGTTAGCAGTTAGACAAATGTTGGAGTGAACCTA 373  
 Db 599 CTGACGATAGGAAAGGAGGAAATGGTTAGGAGGTCGGATAATGTTGGAATGAATTA 658



Db	510	ACTCCTCTTCCACATTTGACTTCAGATGATGTTGATTAAGCCTTACAAAACCTCTCCACGG	569
Qy	705	CARAGATAACAAAGCCAGATTTACCTTATGAGCAAGCGGAGATCAGCGTGAAG	764
Db	570	TTAATGATGTAGAAACACAGATTTACCATATGAGCCCCCAGGAGATCAGCCTGGACC	629
Qy	765	AGTCACAGCCATCCCA---CTCAGTCAAAAGCTACCCAAAGCATCATCTTCAACAGTGCCC	821
Db	630	GGTCAGGSCACCCACGCCCCAGTCGAAAGTGCTCAACCATCTCCTTCCACAGTGCCC	689
Qy	822	AAACAGAAAGCCAGCGTCTCAGTTAGATCCCTTATCAGATTTCTTGACCCGACGACGAGC	881
Db	690	AAAACCTGAAGACCAAGCGCTCTCAGTTAGATCCCTTATCAGATTTCTTGACCCAAAGTAGC	749
Qy	882	CGTCTTGCAAAATCCAGGAGTGGCAGATACAGCTATGGCAGTTCTTACTGGAGCTTCTG	941
Db	750	CGCCTTGCAAAATCCAGCAGTGGCCAGATCCAGCTTTGGCAGTTCTCTCTGGAGCTCCTG	809
Qy	942	TCGACAGCTCCAACTCCAACTGCATCACCTGGGAGGGCAAAATGGGGAGTTCAAGATG	1001
Db	810	TCGACAGCTCCAACTCCAACTGCATCACCTGGGAGGGCAAAACGCGGGAGTTCAAGATG	869
Qy	1002	ACAGACCTGTATGAGTGGCTCGGCGTTGGGAGAGAGAAAGCAACCTTAACATGAAC	1061
Db	870	ACGATCCGAGAGGTGGCCCGCGCTGGGAGAGCGGAGAGCAAAACCAACATGAAC	929
Qy	1062	TATGACAAACTCAGCCGTGCACTTCGCTACTACTATGACAAAATATTATGACTAAAGTT	1121
Db	930	TACGATAAGCTCAGCGCGCCTCCGTTACTACTATGACAAAGAACATCATGACCAAGTC	989
Qy	1122	CATGTTAAACGCTATGCTACAAATTTGATTTCCACGGATCGCTCAGGCGCTCCAGCCT	1181
Db	990	CATGGGAAGCGCTACGCCCTCAAGTTGACTTCCACGGATCGGCCAGGCGCTCCAGCCC	1049
Qy	1182	CACCTCCAGAAATCATCCATGTACAAATPACCCATCAGACCTCCCTACATGAGTTCCCTAC	1241
Db	1050	CACCCCGGAGTCACTCTGTACAGTACCCCTCAGACCTCCCGTACATGGGCTCCTAT	1109
Qy	1242	CATGACACCCCGAGAGATGAACCTTTGTAGCTCCCATCCCTGCTTTGCCCGTAACC	1301
Db	1110	CACGCCACCCACAGAGATGAACCTTTGTGGCGCCCAACCCCTCCAGGCCCTCCCGTGACA	1169
Qy	1302	TCATCCAGCTTTTGTGTCGCCCTAAATCCATCTGGAATTCACCAACTGGAGGCATCTAC	1361
Db	1170	TCTTCCAGTTTTTTGTGTCGCCCAACCCCATCTGGATTCACCACTGGGGGTATATAC	1229
Qy	1362	CCCAATACAGGCTGCCAGCTGCTCATATGCTTCCCATCTTGGCACTTACTACTAA	1418
Db	1230	CCCAACACTAGGCTCCCAACAGGCATATGCTTCTCATCTGGGCACTTACTACTAA	1286

Search completed: May 27, 2004, 19:25:05  
Job time : 5825 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 27, 2004, 16:28:01 ; Search time 674 Seconds  
(without alignments)  
9120.390 Million cell updates/sec

Title: US-09-902-772-1

Perfect score: 1447

Sequence: 1 gaattccgcgaacgaataa.....gaaagaagcgccaagaaaa 1447

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 29Jan04:\*

1: Geneseqn1980s:\*

2: Geneseqn1990s:\*

3: Geneseqn2000s:\*

4: Geneseqn2001as:\*

5: Geneseqn2001bs:\*

6: Geneseqn2002s:\*

7: Geneseqn2003as:\*

8: Geneseqn2003bs:\*

9: Geneseqn2003cs:\*

10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1447	100.0	1447	2	AAx26551 DNA encod
2	1356	93.7	1528	2	AAx26552 DNA encod
3	994.4	68.7	1509	8	Abz68766 Nucleorid
4	921	63.6	1389	7	Abx08759 Angiogene
5	921	63.6	3166	7	Abx76295 Lung canc
6	921	63.6	3166	9	Abd75274 Prostate
7	833.6	57.6	1291	8	Abz68767 Nucleorid
8	742.6	51.3	1372	8	Abz68769 Nucleorid
9	698.8	48.3	1219	8	Abz68768 Nucleorid
10	570.2	39.4	1359	8	Ado2689 Mouse Fli
11	570.2	39.4	1359	9	Adb72427 Mouse Fli
12	570.2	39.4	1729	8	Ado2688 Mouse Fli
13	570.2	39.4	1729	9	Adb72426 Mouse Fli
14	563.8	39.0	2938	2	AAQ50644 Human Hum
15	562.2	38.9	1359	8	Ado2692 Human Fli
16	562.2	38.9	1359	9	Adb72430 Human Fli
17	562.2	38.9	2957	6	Abv94304 Breast ca
18	562.2	38.9	2957	6	Abx84139 Human cDN
19	562.2	38.9	2957	8	Ado2691 Human Fli
20	562.2	38.9	2957	9	Adb72429 Human Fli
21	544.6	37.6	2954	4	AAH02915 Human She
22	397	26.7	567	4	AAI20932 Probe #10
23	387	26.7	567	4	ABA66003 Human foe

C 24	387	26.7	567	4	AAI46176	Aai46176 Probe #14
C 25	387	26.7	567	4	ABA48124	AbA48124 Human bre
C 26	387	26.7	567	4	ABA33090	AbA33090 Probe #11
C 27	387	26.7	567	4	AAK40160	Aak40160 Human bon
C 28	387	26.7	567	4	AAK14426	Aak14426 Human bra
C 29	387	26.7	567	4	ABS39742	AbS39742 Human liv
C 30	387	26.7	567	5	AAI06643	Aai06643 Probe #66
C 31	387	26.7	567	6	ABS14215	AbS14215 Human gen
C 32	358.6	24.8	1431	6	ABS73253	AbS73253 DNA encod
C 33	354.6	24.5	473	4	ABA68128	AbA68128 Human foe
C 34	354.6	24.5	473	4	AAI48339	Aai48339 Probe #17
C 35	354.6	24.5	473	4	ABA50193	AbA50193 Human bre
C 36	354.6	24.5	473	4	ABA35152	AbA35152 Probe #13
C 37	354.6	24.5	473	4	AAK42285	Aak42285 Human bra
C 38	354.6	24.5	473	4	AAK16512	Aak16512 Human bra
C 39	354.6	24.5	473	4	ABS41873	AbS41873 Human liv
C 40	354.6	24.5	473	5	AAI08696	Aai08696 Probe #86
C 41	325	22.5	533	3	AAQ01342	Aaq01342 Human sec
C 42	321.4	22.2	72732	8	ADA02687	Ada02687 Mouse Fli
C 43	321.4	22.2	72732	9	ADb72425	Adb72425 Mouse Fli
C 44	320.4	22.1	1890	2	AAQ50662	Aaq50662 Human Hum
C 45	320.4	22.1	73725	8	ADA02690	Ada02690 Human Fli

## ALIGNMENTS

RESULT 1  
AAx26551  
ID AAx26551 standard; DNA; 1447 BP.  
XX  
AC AAx26551;  
XX  
DT 14-JUN-1999 (first entry)  
XX  
DE DNA encoding chicken C-11 protein.  
XX  
KW Chicken; C-11 protein; cell calcification inhibiting activity;  
KW cell calcification inhibiting agent; c-erg protein; arthritis deformans;  
KW ossification; spinal column ligament; ss.  
XX  
OS Gallus sp.  
XX  
FH Key Location/Qualifiers  
FT CDS 63..1418  
FT /tag= a  
XX  
FN JF11075871-A.  
XX  
PD 23-MAR-1999.  
XX  
PF 29-MAY-1998; 98JP-00166076.  
XX  
PR 18-JUN-1997; 97US-00878177.  
PR 20-JUN-1997; 97US-0050297P.  
XX  
PA (CHUS ) CHUGAI PHARM CO LTD.  
PA (UYPE-) UNIV PENNSYLVANIA.  
XX  
DR WPI: 1999-257708/22.  
DR P-PSDB; AAY01520.  
XX  
FT An active protein for inhibiting cell calcification - useful for  
FT measuring the calcification of a cell, for diagnosing arthritis deformans  
FT or ossification of spinal column ligament.  
XX  
PS Disclosure; Page 7-8; 15pp; Japanese.  
XX  
CC The present sequence encodes a chicken C-11 protein which has cell  
CC calcification inhibiting activity. The specification also describes a  
CC cell calcification inhibiting agent containing c-erg protein (AAV01521).  
CC The proteins are used for measuring the calcification of a cell, for  
CC diagnosing arthritis deformans or ossification of spinal column ligament









angiogenesis-associated disease; cancer; cytostatic; gene therapy; gene; ss.

KW

XX

OS

XX

PN

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

angiogenesis-associated disease; cancer; cytostatic; gene therapy; gene; ss.  
Homo sapiens.  
WO200279492-A2.  
10-OCT-2002.  
14-FEB-2002; 2002WO-US004915.  
14-FEB-2001; 2001US-00784356.  
22-FEB-2001; 2001US-00791390.  
19-APR-2001; 2001US-0285475P.  
03-AUG-2001; 2001US-0310025P.  
13-NOV-2001; 2001US-0350666P.  
29-NOV-2001; 2001US-0334244P.  
(EOSB-) EOS BIOTECHNOLOGY INC.  
Murray R, Glynn R, Watson SR, Aziz N;  
P-PSDB; AB003476.  
WPI; 2003-040681/03.  
P-PSDB; AB003476.

Detecting angiogenesis-associated transcript in a cell for diagnosing and treating cancer by contacting a sample with a polynucleotide that exhibits changes in expression level as a function of time in tissue undergoing angiogenesis.

Example 2; Page 200; 291pp; English.

The present invention relates to methods and compositions for detecting an angiogenesis-associated transcript in a cell in a patient. The method involves contacting a biological sample from the patient with a polynucleotide that selectively hybridises to a sequence at least 80% identical to any of the angiogenesis-associated human polynucleotide sequences given in the specification. These angiogenesis-associated polynucleotide sequences comprise genes that exhibit changes in expression levels as a function of time in tissue undergoing angiogenesis. The method and the polynucleotide sequences of the invention are useful for diagnosing and treating angiogenesis and angiogenesis-associated diseases e.g. cancer. The polynucleotide sequences are also useful in the gene therapy of such disorders. The angiogenesis-associated proteins encoded by the polynucleotide sequences are useful as a vaccine for therapeutic and prophylactic immunisation. ABX08739-BEX08853 represent angiogenesis-associated polynucleotide sequences

Sequence 1389 BP; 359 A; 440 C; 323 G; 267 T; 0 U; 0 Other;

Query Match 63.6%; Score 921; DB 7; Length 1389;  
Best Local Similarity 81.1%; Pred. No. 6.5e-292;  
Matches 1100; Conservative 0; Mismatches 245; Indels 12; Gaps 2;

74 TATTAAGGAAGCATTTACAGTGGTGAAGACAGTCCCTTTGAGTGTGCTACGG 133  
33 TATCAAGGAAGCCTTATCAGTGTGAGTGAAGACAGTGGTGTGCTACGG 92  
134 ATGCCCCACCTTCAAGACAGAAATCAGACCTCTCTCCAGTGAATGGGCAAC 193  
93 AACGCCACCTGGCTAAGACAGAGATGACCGCTCTCTCCAGGACTATGGACAG 152  
194 ATCAAGATGAGCCCGCTTCCAGCAGAGTGGTTATCAGACCCCGCCGACAGT 253  
153 TTCAAGATGAGCCCGCTTCCAGCAGAGTGGTGTCTCAACCCCGACAGGT 212  
254 TACCATTAAGATGAGTGAATCAACCAAGTGAATGGGTCAAGGAATTCACCTGATGA 313  
213 CACCATCAAAATGGAATGAATCAACCTAGCAGGTGAATGGGTCAAGGAATTCCTGATGA 272  
314 CTGACGCGTGGCAAGAGGAGGAAATGGTTAGCAGTTCAGACAAATGTGGGATGAACCTA 373

273 ATGACGTGTGGCCAAAGGGGGAAGATGTGGGACCCAGACACCGCTTGGGATGAACCTA 332  
374 TGGAGCTCATGGAGAGAGACATATTCGGCTCCAAATATGACAAACCAATGAACGAAG 433  
333 CGGACGTACATGGAGAGAGACATGCGCACCCCAACATGACCAACGAGCGCAG 392  
434 AGTTATTGTGCCAGCAGATCCTACGTTATGGAGCAGACCAATGTATCGGAGTGGCTGA 493  
393 AGTTATCGTGGCAGCAGATCCTACGCTATGGAGTACAGACCATGTGCGGAGTGGCTGA 452  
494 GTGGCAGTGAAGGATGTGCTTCCAGACGTGGACATCTTGTGTTCCAGACATTCGA 553  
453 GTGGCGGTGAAAGAAATATGGGCTTCCAGACGTCAACATCTTGTATTCCAGAACATCGA 512  
554 TGGGAAAGAGTGTGTGTAATAATGACAAAGATGACTTCCAGAGACTCACGCCGAGCTATAA 613  
513 TGGGAAAGAACTGTGCAAGATGACCAAGGACGACTTCCAGAGGCTCACCCGAGCTACAA 572  
614 CGCAGATATCTCTCTGTCTACACTTACACTTACACTTACAGAGAGA-----GAGGAGCCAC 664  
573 CGCCGACATCTCTCTCTACACTTCCACTTCCACTTCCAGAGAGACTCTCTTCTTCCACTTTGAC 632  
665 TTTTATTTTCCAAATACATCATAGTTTACCAGAGCAACGCAAGAAATAACCAACAGGCC 724  
633 TTCAGATGATGTGTGATTAAGCCTTACAAACTCTCCAGCTTATGTCATGCTAGAACAC 692  
725 AGATTTACCTTATGACCAAGCAGAGAGATCAGCGTGGACGAGTCAAGCAATCCCA---C 781  
693 AGATTTACCATATGAGCCCCCAGGAGATCAGCGTGGACCGGTACGCGGACACCCACGCC 752  
782 TCAGTCAAAAGCTTACCAACCATCATCTTCAACAGTGGCCCAAAACAGAAAGACAGCGTCC 841  
753 CAGTTCGAAAGCTGTCTCAACCATCTCTTCCAGTGGCCCAAACTGAAGACAGCGTCC 812  
842 TCAGTTAGATCCTTATCAGATCTTGGACCGACAGCGCGTCTTGGAAATCCAGGGAG 901  
813 TCAGTTAGATCCTTATCAGATCTTGGACCAACAGTAGCGCGCTTGCAAATCCAGGCAG 872  
902 TGCGCAGATACAGCTATGSCAGTCTCTACTGAGCTTCTGTGCGACAGTCCAACTCCAA 961  
873 TGCCAGATCCAGCTTTGGCAGTCTCTCTGAGCTCTCTGTGCGACAGTCCAACTCCAG 932  
962 CTGCAATCACCTGGGAGGCGCAAAATGGGAGTTCAAGATGACAGACCTTGATGAAGTGC 1021  
933 CTGCAATCACCTGGGAGGCGCAACGCGGAGTTCAAGATGACGAGTCCCGACGAGTGC 992  
1022 TGGCGTTGGGAGAGAGAAAGCAACCTAACATGAACTATGACAACTCAGCGGTGC 1081  
993 CCGGCGCTGGGAGAGCGGAGCAAAACCCCAACATGAACCTAGATAAGCTCAGCGCGC 1052  
1082 ACTTCGCTACTACTATGACAAAAATATTATGACTAAAGTTCAATGTTAAACGCTATGCCTA 1141  
1053 CCTCGTTACTACTATGACAAAGACATCATGACCAAGGTCCATGGGAGCGCTACGCTTA 1112  
1142 CAAATTTGATTTCCAGGAAATCGCTCAGGCCCTCCAGCTCACCCTCCAGAAATCATCCAT 1201  
1113 CAAATTCGATTTCCAGGGAATCGCCAGGCCCTCCAGGCCCAACCCCGGAGTCTCTCT 1172  
1202 GTACAAATACCATCAGACCTCCCTTACATGATGTTCTTACCATGACACACCCCGAGAGAT 1261  
1173 GTACAGTACCTCTCAGACTCCCGTACATGAGCTCTTATCAGCCCCACCCAGAGAT 1232  
1262 GAATTTGATGTCCTTCCATCCCTTGTGTTGCGGTAACTCATCAGCTTTTGTGCTGC 1321  
1233 GAATTTGTGGGCGCCCAACCTCCAGCCCTCCCGGTGACATCTTCCAGTTTTTGTGCTGC 1292  
1322 CCTAATCCATCTGGAATTCACCAATGAGGAGCTTACCCCAATACAGGCTGCGAGC 1381  
1293 CCAAAACCATCTGGAATTCACCAATGAGGAGCTTATACCCCAACACTAGGCTCCCCAC 1352  
1382 TGTCTATGCTTCCCATCTTGGCACCTTACTACTAA 1418  
1353 CAGCCATATGCTTCTCTCTCTGAGCACTTACTACTAA 1389



Db 1369 CAAGTTGAGCTTCCAGGGATCGCCAGGCGCTCCAGCCACACCCCGGAGTCACTCT 1428  
 Qy 1202 GTACAAATACCCATCAGACTCCCTTACATGAGTTCCTACATGCACACCCCGAGAGAT 1261  
 Db 1429 GTACAAAGTACCCCTCAGACTCCCTTACATGAGTTCCTATCAGCCACCCAGAGAT 1488  
 Qy 1262 GAACCTTTGAGCTCCCATCCCTGCTTTGCGGTAACTCATCCAGCTTTTCTGTC 1321  
 Db 1489 GAACCTTTGCGGCGCCACCTCCAGCCCTCCCGTGACATCTTCCAGTTTTTGTGTC 1548  
 Qy 1322 CCCTAATCCATCTGGAATTCACCACTGGAGGATCTACCCCAATACCAAGGCTGCCAGC 1381  
 Db 1549 CCCAAACCCATCTGGAATTCACCACTGGGGGTATATACCCCACTAGGCTCCCCAC 1608  
 Qy 1382 TGCTCATATGCTTCCATCTTGGCACTTACTACTAA 1418  
 Db 1609 CAGCCATATGCTTCTCATCTGGGCACTTACTACTAA 1645

RESULT 6  
 ID ADB75274 standard; cDNA; 3166 BP.  
 AC ADB75274;  
 XX 04-DEC-2003 (first entry)  
 DE Prostate cancer marker cDNA.  
 KW Prostate; cancer; cytostatic; gene therapy; marker; ss.  
 OS Homo sapiens.  
 XX WO2003009814-A2.  
 XX 06-FEB-2003.  
 XX 25-JUL-2002; 2002WO-US023913.  
 XX 25-JUL-2001; 2001US-0307982P.  
 PR 22-AUG-2001; 2001US-0314356P.  
 PR 25-SEP-2001; 2001US-0325020P.  
 PR 12-DEC-2001; 2001US-0341746P.  
 PR 05-MAR-2002; 2002US-0362158P.  
 XX (MILL-) MILLENNIUM PHARM INC.  
 XX Schlegel R, Monahan JE, Endege WO, Gannavarapu M, Gorbacheva B;  
 PI Hoerish S, Kamatkar S, Wonsey AM, Glatt K, Zhao X, Anderson D;  
 XX WPI; 2003-248033/24.  
 DR New nucleic acid molecule, useful for diagnosing or treating prostate  
 PT cancer.  
 XX Disclosure; SEQ ID NO 98; 99pp; English.

XX The invention relates to newly discovered cancer markers associated with  
 CC the cancerous state of prostate cells. Also disclosed is a method of  
 CC assessing whether a patient is afflicted with prostate cancer. The method  
 CC of the invention involves assessing whether a patient is afflicted with  
 CC prostate cancer by comparing the level of expression of a marker in a  
 CC patient sample and the normal level of expression of the marker in a  
 CC control non-prostate cancer sample, where a significant increase in the  
 CC level of expression of the marker in the patient sample and the normal  
 CC level indicates that the patient is afflicted with prostate cancer.  
 CC Nucleic acids of the invention are useful for diagnosing or treating  
 CC prostate cancer, and may be useful in gene therapy. Sequences given in  
 CC AD875177-ADB75631 represent marker cDNA and proteins. Note: The sequence  
 CC data for this patent did not form part of the printed specification, but  
 CC was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pat\_sequences.

XX SQ Sequence 3166 BP; 915 A; 770 C; 768 G; 713 T; 0 U; 0 Other;  
 Query Match 63.6%; Score 921; DB 9; Length 3166;  
 Best Local Similarity 81.1%; Pred. No. 1.1e-291;  
 Matches 1100; Conservative 0; Mismatches 245; Indels 12; Gaps 2;  
 Qy 74 TATTAAAGGAAGCATATCAGTGGTGAAGACAGCCAGTCTTGTGAGTGTGCTACGG 133  
 Db 289 TATCAAGGAGCCTTATCAGTTGTGAGTGAAGACAGTCTTGTGAGTGTGCTACGG 348  
 Qy 134 ATCGCCCACTTGAAGAGACAGAAATGACAGCTCTCTTCCAGTGAATATGGCAAAAC 193  
 Db 349 AAGCCACACCTGGCTAAGACAGAGATGACCGCGTCTCTCCAGCGACTATGGACAGAC 408  
 Qy 194 ATCAAGATGAGCCCGCGGTTCACAGAGAGACTGGTATCAAGAGCCCGCCGAGGT 253  
 Db 409 TTTCAAGATGAGCCCGCGTCTCTCAAGAGATGGTGTCTCAACCCCGCAGCGGT 468  
 Qy 254 TACCATTAAAGATGAGTGTAAACCCAAACAGCTTAATGGGTCAAGAAATTCACCTGATGA 313  
 Db 469 CACCATCAAAATGGAATGTAAACCTAGCCAGGTGAATGGTCAAGAAATCTCTCTGATGA 528  
 Qy 314 CTGAGCGTGGCAAAAGGAGGAAATGTTAGCAGTTTCAAGCAATGTTGGATGAACCTA 373  
 Db 529 ATSCAGTGTGGCCAAAGCGGAGAGTGTGGGACGCCAGACACCTGTTGGATGAACCTA 588  
 Qy 374 TGGAAAGTACATGGAAGAGAGAGCATATTCCGCTCCAAATATGACAAACCAATGAACGAAG 433  
 Db 589 CGGAGCTACATGGAAGAGAGAGCATATGCCACCCCAACATGACCAAGAGCGGAG 648  
 Qy 434 AGTTATGTGCGCAGAGATCTTACGTTATGAGAGCAGACCAATGACGAGTGGCTGGA 493  
 Db 649 AGTTATGTGCGCAGAGATCTTACGCTATGGAGTACAGACCATGTGCGGCGAGTGGCTGGA 708  
 Qy 494 GTGGCAGTGAAGAGTATGGTCTTCCAGAGCTGACATCTTGTGTTTCCAGAACTGA 553  
 Db 709 GTGGCAGTGAAGAAATATGGCTTCCAGAGCTCAACATCTTGTATTCAGAACTGA 768  
 Qy 554 TGGAAAGAGTGTGTAAATGACCAAGATGACTTCCAGAGACTCAGCCGAGCTATAA 613  
 Db 769 TGGAAAGGAACCTGTGCAAGATGACCAAGAGAGCTTCCAGAGGCTCAGCCCGAGTACAA 828  
 Qy 614 CGCAGATATCTCTGTGTCACACTACTACTCTCAGAGAGA-----GAGGAGCCAC 664  
 Db 829 CGCCGACACTCTTCTCAGACTCTCCACTCTCAGAGAGACTCTCTTCCAGATTTGAC 888  
 Qy 665 TTTTATTTTCCAAATACATCAGTTTACCCAGAGCAAGCAAGCAAGATAACCAAGGCC 724  
 Db 889 TTCAGATGATGTTGATAAAGCCTTACAAACCTCTCCAGGTTAATGCATGCTAGAAACAC 948  
 Qy 725 AGATTACCTTATGAGCAAGCGAGAGATCAGCGTGGAGAGTCAAGCCATCCCA---C 781  
 Db 949 AGATTACCATATGAGCCCCCAGAGATCAGCGTGGAGCGGTACGGCCACCCACGCC 1008  
 Qy 782 TCAGTCAAAAGCTAGCCCAACCATCATCTTCAACAGTGGCCCAAAACAGAGACAGCGTCC 841  
 Db 1009 CAGTCAAGAGCTGTCAACCATCTCTTCCAGTGGCCCAAACTGAAGACAGCGTCC 1068  
 Qy 842 TCAGTTAGATCTTATCAGATCTTGGACCGGACAGAGCGCTCTTGCMAATCCAGGAG 901  
 Db 1069 TCAGTTAGATCTTATCAGATCTTGGACCAACAGATAGCGCGCTTGCMAATCCAGGAG 1128  
 Qy 902 TGGGAGATACAGCTATGGCAGTCTCTACTGAGGCTTCTGTCGGAGAGCTCCCAATCCAA 961  
 Db 1129 TGGCAGATCCAGCTTTGGCAGTTCTCTCTGAGCTCTCTGTCGGAGAGTCCCACTCCAG 1188  
 Qy 962 CTGCATACCTGGGAGGCGCAAAATGGGAGTTCAGATGACAGACCTCTGATGAAGTGGC 1021  
 Db 1189 CTGCATACCTGGGAGGCGCAACCGGAGTTCAGATGACAGGATCCCGAGAGTGGC 1248  
 Qy 1022 TGGCGTGGGAGAGAGAAAGCAACCTAACATGACTATGACAACTCAGCGGTCC 1081



QY 1071 CTCAGCGTGCACCTTCGCTACTACTATGACAAAATATTATGACTAAAGTTTCATGGTAAA 1130  
 DB 881 CTCAGCGCGCCCTCCGTTACTACTATGACAAAGAACATCATGACCAAGGTCATGGGAAG 940  
 QY 1131 CGCTATGCTACAAATTTGATTTCCACGGAATCGCTCAGGCCCTCCAGCCTCACCTCCA 1190  
 DB 941 CGCTAGGCTACAGATTCGATTTCCACGGGATCGCCAGGCCCTCCAGCCTCACCTCCA 1000  
 QY 1191 GAATCATCCATGTAACAAATACCATCAGACCTCCCTACATGATTTCTTACCATGCAAC 1250  
 DB 1001 GAGTCATCTCTGTACAGTACCCCTCAGACCTCCCTACATGATGCTTCTATCAGGCCAC 1060  
 QY 1251 CCCAGAGAGATGAATTTGTAGTCCCTATCCCTCCCTGTTTGGCCGTAACCTCATCCAGC 1310  
 DB 1061 CCACAGAGATGAATTTGTGGCGCCCACTCCAGCCCTCCCGTGACATTTCCAGT 1120  
 QY 1311 TTTTGTGTCCTTAATCCATCTACTGGAATTCACCAAGTGGAGCATCTACCCCAATACC 1370  
 DB 1121 TTTTGTGTCCTTAATCCATCTACTGGAATTCACCAAGTGGGATATATACCCCAACT 1180  
 QY 1371 AGGCTGCCAGTGTCTCATATGCTTCCCTTCCCTTGGCCTACTACTAA 1418  
 DB 1181 AGGCTCCCAAGCAGCATATGCTTCTCATCTGGGCATTTACTACTAA 1228

RESULT 8  
 ABZ68769  
 ID ABZ68769 standard; DNA; 1372 BP.  
 XX  
 AC ABZ68769;  
 XX  
 DT 16-MAY-2003 (first entry)  
 XX  
 DE Nucleotide sequence of human ERG splice variant C-1-4.  
 XX  
 KW Human; splice variant; ERG; Ets; transcription factor; C-1-1; C-1-2;  
 KW C-1-3; C-1-4; chondrocyte; articular chondrocyte; bone disease;  
 KW cartilage disease; tumour; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 146..1309  
 FT /\*tag= a  
 FT /product= "ERG splice variant C-1-4"  
 XX  
 PN WO2003000724-A2.  
 XX  
 PD 03-JAN-2003.  
 XX  
 PF 08-MAR-2002; 2002WO-EP002605.  
 XX  
 PR 25-JUN-2001; 2001US-0300756P.  
 XX  
 PA (TIGER-) TIGENIX NV.  
 XX  
 PI Luyten F, De Bari C, Dell'accio F;  
 XX  
 DR WPI; 2003-221398/21.  
 DR P-PSDB; ABP97698.  
 XX  
 PT Novel splice variants of the Ets-related gene (ERG) transcription factor  
 PT , useful for preventing or treating bone-related or cartilage-related  
 PT diseases or Ets-related gene associated tumor.  
 XX  
 PS Claim 112; Fig 7; 60pp; English.  
 XX  
 CC The present sequence encodes splice variant C-1-4 of human ERG. ERG  
 CC belongs to the Ets family of transcription factors. The specification  
 CC describes C-1-1 (characterised by the absence of the 81 bp fragment, and  
 CC the presence of the 218 bp and 72 bp fragments of the ERG-3 cDNA  
 CC sequence); C-1-2 (characterised by the absence of the 81 bp and 218 bp  
 CC fragments, and the presence of the 72 bp fragment of the ERG-3 cDNA

CC sequence); C-1-3 (characterised by the absence of the 72 bp, 81 bp and  
 CC 218 bp fragments of the ERG-3 cDNA sequence); and C-1-4 (characterised by  
 CC the absence of the 218 bp fragment, and the presence of the 72 bp and 81  
 CC bp fragments of the ERG-3 cDNA sequence). The splice variant  
 CC polynucleotides are useful for stabilizing the phenotype of chondrocytes  
 CC or cells with a chondrocyte-like phenotype, for preventing de-  
 CC differentiation of articular chondrocytes during in vitro culture  
 CC expansion, for preventing or treating bone-related or cartilage-related  
 CC diseases or Ets-related gene associated tumour, or as marker of  
 CC progenitor cells of stable chondrocytes. Cells expressing the  
 CC polynucleotide are useful for producing or repairing cartilage tissue in  
 CC a mammal  
 XX  
 SQ Sequence 1372 BP; 359 A; 430 C; 314 G; 269 T; 0 U; 0 Other;  
 Query Match 51.3%; Score 742.6; DB 8; Length 1372;  
 Best Local Similarity 78.7%; Pred. No. 3.8e-233;  
 Matches 951; Conservative 0; Mismatches 174; Indels 84; Gaps 2;  
 QY 294 TCAGGAATTCACCTGATGCTCAGCGTGGCAAGAGGAGGAAATGTTAGCAGTTCA 353  
 DB 101 TCAGGAATTCCTCTGATGAATGCAAGTGGCCAAAGGCGGAAAGATGTGGCAGCCCA 160  
 QY 354 GACAATGTTGGATGAATATGGAAGCTACATGGAAGAGCAATATCCGGCTCCAAAT 413  
 DB 161 GACACCGTTGGATGAATACCGCAGCTACATGAGGAGGAGCAGATGCCCAAAAC 220  
 QY 414 ATGACAAACCAATGACGAGAGGTATTGTCAGCAGAGATCCTACGTTATGGAGCAGAC 473  
 DB 221 ATGACCAACAGCAGCGCAGAGTTATGTCGACAGAGATCTTACGCTATGGAGTACAGAC 280  
 QY 474 CATGTACGGCAGTGGCTGGAGTGGCGAGTGAAGAGTATGTCCTTCCAGACGTGGACATC 533  
 DB 281 CATGTGGCGAGTGGCTGGAGTGGCGGTGAAAGAAATATGGCTTCCAGACGTCAACATC 340  
 QY 534 TTGTTGTTCCAGAACATTTGATGGGAAAGAGTTGTGTAATGACCAAGATGACTTCCAG 593  
 DB 341 TTGTTATTCAGAACATTCGATGGGAAAGAACTGTCCAGATGACCAAGGAGCAGTTCCAG 400  
 QY 594 AGACTCACGCCGAGCTATTAACGAGATATCCTCTGTGTACACCTACACTACCTCAGAGAG 853  
 DB 401 AGGCTCACCCCGAGCTACAAACGCCGACATCCTTCTCTACATCTCCACTACCTCAGAGAG 460  
 QY 654 A----- 654  
 DB 461 ACTCCTCTTCCACATTTGACTTTCAGATGATGTTGTAAGCCCTTACAAACTCTCCACGG 520  
 QY 655 -----GAGGAGCCACTTTTATTTTCCAAATACATCAGTTTAC 692  
 DB 521 TTAATGATGCTAGAAACACAGGGGTGCAGCTTTTATTTTCCCAATATCTCAGTATAT 580  
 QY 693 CCAGAGCAACGCAAGAAATAACAAGCCAGCCAGATTTACCTTATGAGCAAGCGAGAGA 752  
 DB 581 CCTGAAGCTACGCAAGAAATTACAACTAGGCCAGATTTACCATATGAGCCGCCAGAGA 640  
 QY 753 TCAGCGTGCAGAGTACAGCCATCCCTCT---CAGTCAAAAGCTACCCCAATCATCT 809  
 DB 641 TCAGCCTGCAGCCGTCACGCCACCCCGCCAGTCGAAAGCTGCTCAACCATCTCT 700  
 QY 810 TCAACAGTGCCTCAAAACACAGAACACAGCGCTCCTCTAGTTAGATCTTATCAGATTCTGA 869  
 DB 701 TCCACAGTGCCTCAAAACACAGAACACAGCGCTCCTCTAGTTAGATCTTATCAGATTCTGA 760  
 QY 870 CCGACAGCAGCGCTCTTCCAAATCCAGGAGTGGCAGATACAGTATGGCAGTTCTTA 929  
 DB 761 CCAACAAAGTAGCCGCTTCCAAATCCAGGAGTGGCCAGATCCAGGTTTGGCAGTTTCTC 820  
 QY 930 CTGAGCTTCTGTGCGACAGCTCCAACTCAACTGCATCCTCTGGGAGGCGCAAAATGGG 989  
 DB 821 CTGGAGCTCTGTGCGACAGCTCCAACTCCAGTGCATCCTCTGGGAGGCGCAAAACGGG 880  
 QY 990 GAGTTCAAGATGACAGCCCTGATGAAGTGGCTCGCGGTTGGGAGAGAGGAAAGCAAA 1049







QY 1011 GATGAAGTGGCTCGCGTGTGGGAGAGAGAAAGCAAACTAATGAACTATGACAAA 1070  
Db |||||  
749 GACGAGGTGGCCGCGCTGGGAGAGCGGAGAGCAACCACTGAATGATAAG 808  
QY 1071 CTCAGCGTGCACCTCGCTACTACTATGACAAAATATATGACTAAAGTTGATGTA 1130  
Db |||||  
809 CTCAGCGCGCGCTCGCTACTACTATGACAAAATATATGACTAAAGTTGATGTA 868  
QY 1131 CGCTATGCGCTCAAAATTTGATTTTCCAGCAATCGCTCAGGCGCTCCAGCGCTCACCTCCA 1190  
Db |||||  
869 CGCTACGCGCTCAAAATTTGATTTTCCAGCAATCGCTCAGGCGCTCCAGCGCTCACCTCCA 928  
QY 1191 GAATCATCATGACAAATACCATGACACTCCCTCATGAGTTCTTACCATGACAC 1250  
Db |||||  
929 GAGTCATCTGTCAAGTACCCCTCAGACCTCCCGTACATGCGCTCTCTATCAGCGCCAC 988  
QY 1251 CCCAGAGAGTGAATTTGTAGCTCCCATCCCTGCTTTTGGCCGTAACCTCATCCAGC 1310  
Db |||||  
989 CCACAGAGATGAATTTGTGGCGCCACCCCTCAGCGCTCCCGTGACATCTTCCAGT 1048  
QY 1311 TTTTGTGTCGCCCTTAATCCATCTACTGGAATTCACAACTGAGGCACTTACCCCAATACC 1370  
Db |||||  
1049 TTTTGTGTCGCCCAACCCATCTGGAATTCACAACTGAGGCACTTATACCCCAACT 1108  
QY 1371 AGGCTGCCAGTGTCTCATGCTCCCTTCCATCTTGGCACCTACTACTAA 1418  
Db |||||  
1109 AGGCTCCCAACAGCCATATGCTTCTCATCTGGGCACTTACTACTAA 1156

RESULT 10  
ADA02689  
ID ADA02689 standard; cDNA; 1359 BP.  
XX  
AC ADA02689;  
XX  
DT 06-NOV-2003 (first entry)  
XX  
DE Mouse Flil carcinoma associated coding sequence, SEQ ID NO:1207.  
XX  
KW Mouse; murine; carcinoma associated; oncogene; carcinoma; cancer; breast;  
KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;  
KW gene; ss.  
XX  
OS Mus sp.  
XX  
PN WO2003057146-A2.  
XX  
PD 17-JUL-2003.  
XX  
PF 26-DEC-2002; 2002WO-US041414.  
XX  
PR 26-DEC-2001; 2001US-00035832.  
XX  
PA (SAGR-) SAGRES DISCOVERY.  
XX  
PI Morris DW;  
XX  
DR WPI; 2003-587068/55.  
XX  
PT New recombinant nucleic acid encoding carcinoma associated protein,  
PT useful for preparing compositions for treating carcinomas.  
XX  
PS Claim 1; SEQ ID NO 1207; 245pp; English.  
XX  
CC The invention relates to recombinant carcinoma associated (CA) nucleic  
CC acid sequences from mouse and human (ADA01482-ADA03094), and to  
CC recombinant carcinoma associated proteins (CAP) encoded by them. The  
CC invention also encompasses expression vectors and host cells comprising a  
CC CA nucleic acid, a polypeptide (especially an antibody) that specifically  
CC binds to the protein, and a biochip comprising CA nucleic acid or  
CC fragments thereof. The sequences of the invention were identified using  
CC oncogenic retroviruses, which insert into the genome of the host organism  
CC at random. Many of these do not carry transduced host oncogenes or

CC pathogenic trans-acting viral genes, meaning that cancer incidence is a  
CC direct consequence of the effects of proviral integration into host  
CC protooncogenes. The CA nucleic acid sequences can be used to diagnose  
CC carcinoma (especially breast cancer, prostate cancer, lymphoma or  
CC leukaemia) or a propensity to carcinoma by determination of the sequence  
CC of a CA gene, or by determination of CA gene expression in particular  
CC tissues. CA nucleic acids, proteins and antibodies are also useful as  
CC therapeutic agents and in screening and evaluating drug candidates. The  
CC present sequence represents a specifically claimed murine CA nucleic acid  
CC sequence of the invention. Note: The complete sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 1359 BP; 367 A; 398 C; 331 G; 263 T; 0 U; 0 Other;  
Query Match 39.4%; Score 570.2; DB 8; Length 1359;  
Best Local Similarity 66.8%; Pred. No. 2.4e-176;  
Matches 917; Conservative 0; Mismatches 423; Indels 33; Gaps 6;  
QY 63 ATGGCAAGCACTATTAAAGGAGCATTTATCAGTGTGAGTGAAGACAGCTCTTTGTTGAG 122  
Db |||||  
1 1ATGGACGGGACTATTAAAGGAGGCTCTGTCTGTGGTGAGTGACATCAGTCCCTTTTGTAT 60  
QY 123 TGTGCTACGGATCG---CCCCACCTTGCAGAGAGAAATCAGACAGCTCTCTTCCAGT 179  
Db |||||  
61 TCAGCATACGGAGCGCGACCCATCTCCCAAGGAGATATGACTGCTTCGGGAGTCTCT 120  
QY 180 GAATATGGGCAAAACATCAAGATGAGCCCGCGCTTCCCGACGAGCTGGTATTACAG 239  
Db |||||  
121 GACTACGGGCGAGCCCGACAAATCAACCCCTGCCACGCGAGGAGTGGATCAACCG 180  
QY 240 CCCCOCGCCAGAGTTACCATTAAGATGAGTGTAAACCCAAACAGAGTTAATGGTCAAGG 299  
Db |||||  
181 CC-----AGTGAGAGTCAATGTCAAGCGGGAGTATGACCACATGAATGATCCAGG 231  
QY 300 AATTACCTGATGACTGCGAGCGTGGCAAGAGGAGGAAATGTTAGCAGTTTCAGACAAT 359  
Db |||||  
232 GAGTCTCCGGTGCAGTGTGAGTGTGACAAATGTAACAGCTGTTGGCGGAGCGAGCC 291  
QY 360 GTTGGATGAATATGAGAGCTACATGGAAGAGAGCA---TATTCCGCTTCAATATG 416  
Db |||||  
292 AATCCCATGAATATAATAGTACATGATGATGAGAGACGCGCCCTCTCTCCCAATG 351  
QY 417 ACAACCAATGAAGAGAGTATTGTGCCAGCAGATCTTACGTTATGAGGACAGACCAT 476  
Db |||||  
352 ACCACCAAGAGAGAGTCAATGTGCTGCGAGCCCGACACTGTGACAGAGGAC 411  
QY 477 GTACGGCAGTGTGGAGTGGCAGTGAAGAGTATGTTTCCAGACGTGGACATCTTTG 536  
Db |||||  
412 GTTCGACAGTGTGGAGTGGGCTATAAGGAATACGGGATTGATGGAGATTGACACTTCC 471  
QY 537 TTGTTCCAGACATTTGATGGGAGAGTGTGTAATATGACCAAGATGACTTCCAGAGA 596  
Db |||||  
472 TTCTTCCAGACATGATGGAGAGGAAATTTGTAAATATGAACAGAGGAGACTTCTCCGA 531  
QY 597 CTCACGCCGAGCTATAACGCGAGATATCTCTCTGTCTCACACTACACTACCTCAGAGAGA 656  
Db |||||  
532 GCCACCTCCGCTACACACAGAGTGTGTTGTGCGACCTCAGTTACCTCAGGGAAG- 590  
QY 657 GGAGCCACTTTATTTTCCAAATACATCAGTTTCCCGAGAGCAACGGAAGATAACA 716  
Db |||||  
591 -----TTCACTGCTGGCCTTATACCAACCTCCCATACAGACCGTCTTCAGACTGAAT 645  
QY 717 ACAAGGCCAGATTACCTTTATGAGCAAGCGAGAGATCAGCGTGGACGAGTACACA---GC 773  
Db |||||  
646 GTCAAGAGAGACCTTCTTATGACTGTGTGAGGAGAGGAGCATGGACATATATGAAC 705  
QY 774 CATCCCACTCAGTCAAAAGCTACCCAAACCATCATCTTCAACAGTGGCCCAAAACAGAGAC 833  
Db |||||  
706 TCTGGCCTCAACAAAAGTCTCTCTCTTGGAGGATCAGACCATGGGCAAGAACACTGAG 765  
QY 834 CAGCGTCTCAGTTAGATCTTATCAGATTTCTTGGACCGGACGAGCGCTCTTGCAAT 893



Db 766 CAGCGGCCCCAGACGATCCTTATCAGATCTCTGGGGCCAAACGAGCGCGCTAGCAAC 825  
QY 894 CCAGGAGTGGCGAGATACAGTATGGAGTCTCTACTGGAGTCTTCTGCGGACAGCTCC 953  
Db 826 CTTGGAGTGGCGAGATCCAGCTGTGGAGTTCTCTGGAACTACTCTCGACAGCGCC 885  
QY 954 AACTCCAACTGCATCACCTGGGAGGCGCAAAATGGGAGTTTCAAGATGACAGACCTGAT 1013  
Db 886 AACGCCAGCTGTATCACCTGGGAGGAGCCAAACGGGAGTTTCAAAATGACGACCGCTGAT 945  
QY 1014 GAAGTGGCTGGCGTGGGAGAGAGAAAGCAAACTATACATGAATATGACAAATC 1073  
Db 946 GAGGTGGCCAGCGCTGGGAGAGCGAAGAGCAAGCCCAACATGAATATGACAAGCTG 1005  
QY 1074 ACCCTGCACTTCGCTACTACTATACAAAAATATATGATTAAGTTTCATGTTAAACGC 1133  
Db 1006 AGCCGGGCCCCGATACTACTATGACAAAACATATATGACAAAGTGCATGGCAAGG 1065  
QY 1134 TATGCTCAAAATTTGATTTTCCAGGAATCGCTCAGGCGCTCCAGCTTCACCTCCAGAA 1193  
Db 1066 TATGCTCAAAATTTGATTTTCCAGGAATCGCTCAGGCGCTCCAGCTTCACCTCCAGAA 1125  
QY 1194 TCATCCATGTAACAATCCCATCAGACCTCCCTACATGACTTCTACCATGCAACCCC 1253  
Db 1126 ACATCCATGTAACAATCCCTCTGATATCTCTACATGCTTCTACCATGCCCATCAA 1185  
QY 1254 CAGAAAGATGAATTTGTAGTCTCCCATCCCTGCTTTTGGCCGTAACCTCATCCAGCTTT 1313  
Db 1186 CAGAAAGTGAATTTGTCCCGTCTCAACCATCTCTCCATGCTGTACCTCTCCAGCTTC 1245  
QY 1314 TTTGTGCCCCATATCCATATCTGGAATTCACCAACTGGAGGCACTACCCCAATACCGG 1373  
Db 1246 TTTGGAGCAGCATCAATATCTGGAACCTCCCTCCCTGCTGGATCTATCCAAACCCAGT 1305  
QY 1374 CTGCCAGCTGCTCATA-----TGCTTCCCATCTTGGCACTACTACTA 1417  
Db 1306 GTCCCCCGCCATCTTAACACCCACGTGCTTACACTTAGGCGAGTACTACTA 1358

## RESULT 12

ID ADA02688 standard; cdNA; 1729 BP.  
XX ADA02688;  
AC ADA02688;  
XX  
DT 06-NOV-2003 (first entry)  
XX  
DE Mouse Flil carcinoma associated cdNA, SEQ ID NO:1206.  
XX  
KW Mouse; murine; carcinoma associated; oncogene; carcinoma; cancer; breast;  
KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;  
KW gene; ss.  
XX  
OS Mus sp.  
XX  
PN WO2003057146-A2.  
XX  
PD 17-JUL-2003.  
XX  
PF 26-DEC-2002; 2002WO-US041414.  
XX  
PR 26-DEC-2001; 2001US-00035832.  
XX  
PA (SAGR-) SAGRES DISCOVERY.  
XX  
PI Morris DW;  
XX  
DR WPI; 2003-587068/55.  
XX  
PT New recombinant nucleic acid encoding carcinoma associated protein,  
PT useful for preparing compositions for treating carcinomas.  
XX  
PS Claim 1; SEQ ID NO 1206; 245pp; English.

XX The invention relates to recombinant carcinoma associated (CA) nucleic  
CC acid sequences from mouse and human (ADA01482-ADA03054), and to  
CC recombinant carcinoma associated proteins (CAP) encoded by them. The  
CC invention also encompasses expression vectors and host cells comprising a  
CC CA nucleic acid, a polypeptide (especially an antibody) that specifically  
CC binds to the protein, and a biochip comprising CA nucleic acid or  
CC fragments thereof. The sequences of the invention were identified using  
CC oncogenic retroviruses, which insert into the genome of the host organism  
CC at random. Many of these do not carry transduced host oncogenes or  
CC pathogenic trans-acting viral genes, meaning that cancer incidence is a  
CC direct consequence of the effects of proviral integration into host  
CC protooncogenes. The CA nucleic acid sequences can be used to diagnose  
CC carcinoma (especially breast cancer, prostate cancer, lymphoma or  
CC leukaemia) or a propensity to carcinoma by determination of the sequence  
CC of a CA gene, or by determination of CA gene expression in particular  
CC tissues. CA nucleic acids, proteins and antibodies are also useful as  
CC therapeutic agents and in screening and evaluating drug candidates. The  
CC present sequence represents a specifically claimed murine CA nucleic acid  
CC sequence of the invention. Note: The complete sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX

SEQ Sequence 1729 BP; 466 A; 484 C; 436 G; 343 T; 0 U; 0 Other;

Query Match 39.4%; Score 570.2; DB 8; Length 1729;  
Best Local Similarity 66.8%; Pred. No. 2.8e-176;  
Matches 917; Conservative 0; Mismatches 423; Indels 33; Gaps 6;  
QY 63 ATGGCAAGCACTATTAAAGAGCATATCAGTGGTGAAGACCAAGCTTGTGTTGAG 122  
Db 245 ATGGAGCGGCACTATTAAAGAGGCTCTGTGTGTGAGTGAAGTCAAGTCACTCTTTTGTAT 304  
QY 123 TGTGCTCTACGATCG--CCACACCTTGCAAGACAGAAATGACAGCTCTCTCCAGT 179  
Db 305 TCAGCATACGGAGCGCAGCCCATCTCCCAAGGAGAGATGACTGCTTCGGGGAGTCT 364  
QY 180 GAATATGGGCAACATCAAGATGAGCCGCGGTTCCTCCAGCAGAGCTGTTATCACAG 239  
Db 365 GACTACGGGCGAGCCGCCAAATCAACCCCTGTCACCGCAGCAGAGTGAATCAACAC 424  
QY 240 CCCCAGCGCAGAGTTACCATTAAGATGGAGTGTAAACCAACAGCTTAAATGGTCAAGG 299  
Db 425 CC-----AGTGAGATCATGTCAAGCGGAGTATGACCAATGATGATCCAGG 475  
QY 300 AATTCACTGATGACTGCGCGGTGGCAAAAGGAGGAAATAGTTAGAGTTCAGACAAAT 359  
Db 476 GAGTCTCCGTTGACTGCACTGTCAAGAAATGTAAACAGCTGTGTGGGCGGAGCGAAGCC 535  
QY 360 GTTGGGATGAACATATGGAAGCTACATGGAAGAGAGCA--TATTCCGCTCCCAATATG 416  
Db 536 AACCCATGAACCTATTAATAGTACATGATGAGAGAGAGAGCA--TATTCCGCTCCCAATATG 595  
QY 417 ACAACCAATGAACGAAGAGTATTGTGCCAGCAGATCTTACGTTATGAGAGCAGACCAT 476  
Db 596 ACCACCAACGAACGGAGAGTCAATTTGTCTGACAGACCCCACTGTGGACACAGGAGCAC 655  
QY 477 GTACGGAGTGGCTGAGTGGGCGAGTGAAGAGTATGGTCTCCAGACGCTGGAGTCTTG 536  
Db 656 GTTCGACAGTGGCTGAGTGGGCTATAAAGGAATACGGATTAATGAGAGATTGACATTC 715  
QY 537 TTGTTCCAGAACTTGTATGGGAAAGAGTGTGTAAATGACCAAGATGACTTCCAGAGA 596  
Db 716 TTCTTCCAGACATGATGGCAGGAAATTTGTAAATGACCAAGAGGAGTCTTCTCCGA 775  
QY 597 CTCAGCGGAGCTATTAACGAGATATCTCTCTGTCTACACTACCTACCTCAGAGAGA 656  
Db 776 GCCACCTCCGCTTACCAACAGAAAGTGTGTGTGCGACCTCAGTTTACCTCAGGGAAG- 834  
QY 657 GGAGGCCACTTTTATTTTCCAAATACATCAGTTTACCCAGAGCAACCGCAAGAAATAACA 716  
Db 835 -----TTCAGTGTGGCTTATAACAACTCCCATACAGACCGTCTCTCACAGTGAAT 889

```

QY 717 ACAAGGCCAGATTTACCTTATGACGACGAGGAGATCAGCGTGGAGAGTCACA---GC 773
Db 890 GTCAGGAGAGACCCCTTCTATGACTCTCTCAGGAGAGGAGCATGGAACATAATATGAAC 949
QY 774 CATCCACTAGTCACAAAGCTACCAACCATCATCTTCAACAGTSCCCAAACAGAGAC 833
Db 950 TCTGGCCCTCAACAAAGTCTCTCTCTGGAGGATCAGACCATGGCAAGAACACTGAG 1009
QY 834 CAGCGTCCCTCAGTTAGATCCCTTATCAGATTTCTGGACCGACGACGCGCTCTGCAAT 893
Db 1010 CAGCGGCCCGGACGAGATCCCTTATCAGATCTCTGGGCGAACACGACGCGCTTAGCAAC 1069
QY 894 CAGGAGTGGGCGAGATACAGTATGGCAGTTCCTTACTGAGCTTCTGTGGACAGCTCC 953
Db 1070 CCTGGAGTGGGCGAGATCCAGCTGTGGCAGTTCTCTGGAACACTACTGTCGACAGCGC 1129
QY 954 AACTCCACTGCTACCTCTGGAGGACAAATGGGAGTTCAGATGACAGACCTTGAT 1013
Db 1130 AACCGCAGCTATCCTCTGGAGGGGACCAACGGGAGTTCAAAATGACGGACCTGAT 1189
QY 1014 GAAAGTGGCTCGCGTGGGAGAGAGGAAAGCAAACTAACATGAATGACAAACTC 1073
Db 1190 GAGGTGGCCAGCGCTGGGAGAGCGGAGAGCAAGCCCAACATGATTATGACAGCTG 1249
QY 1074 AGCGTGCTCAGTCTGCTACTATGACGACAAATATGACTAAAGTTCTAGTAAACGC 1133
Db 1250 AGCGGGGCCCTCCGATACTACTATGACAAACAAATATGACAAAGTGCATGGCAAGG 1309
QY 1134 TATGCTCAAAATTTGATTTCCAGGAATGCTCAGGCGCTCCAGCGCTCACCCTCCAGAA 1193
Db 1310 TATGCTCAAAATTTGATTTCCAGGAATGCTCAGGCGCTCCAGCGCTCACCCTCCAGAA 1369
QY 1194 TCATCCATGTACAAATACCCATGACAGCTCCCTTACATGAGTTCCTACCATGACACCCC 1253
Db 1370 ACATCCATGTACAAATACCCATGACAGCTCCCTTACATGAGTTCCTACCATGACACCCC 1429
QY 1254 CAGAGATGAACTTTGTAGCTCCCATCCCTCTGTTTGGCCGTAACCTCATCCAGCTTT 1313
Db 1430 CAGAGATGAACTTTGTAGCTCCCATCCCTCTGTTTGGCCGTAACCTCATCCAGCTTT 1489
QY 1314 TTTGCTGCCCTTAATCCATCTGGAATTCACCAACTGGAGCATCTACCCCAATACCAAG 1373
Db 1490 TTTGGAGCAGCATACAAATCTGGACCTCCCTCCCTGCTGGGATCTATCCAAACCCCACT 1549
QY 1374 CTCGACGCTGCTCAT------TGCCTTCCCATCTTGGCAGCTACTACTA 1417
Db 1550 GTCCCCCGCCATCCTTAACACCCACGTCGCTTTCACACTTAGGACGCTACTACTA 1602

```

```

RESULT 13
ADB72426
ID ADB72426 standard; mRNA; 1729 BP.
XX AC ADB72426;
XX DT 04-DEC-2003 (first entry)
XX DE Mouse Flil mRNA.
XX KW mouse; ss; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
XX LW cancer; neoplasm; adenocarcinoma; sarcoma.
XX OS Mus sp.
XX PN WO2003008583-A2.
XX PD 30-JAN-2003.
XX PF 26-DEC-2001; 2001WO-US051291.
XX PR 02-MAR-2001; 2001US-00798586.
XX PR 23-OCT-2001; 2001US-00004113.

```

```

PR 08-NOV-2001; 2001US-00052482.
PR 30-NOV-2001; 2001US-00997722.
PR 20-DEC-2001; 2001US-00034650.
XX (SAGR-) SAGRES DISCOVERY.
PA Morris DW, Engelhard EK;
PI WPI; 2003-239337/23.
XX New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
PT cancers, neoplasm, adenocarcinoma, or sarcomas.
XX Claim 1; SEQ ID NO 254; 2304pp; English.
PS The invention relates to a novel recombinant nucleic acid comprising a
CC nucleotide sequence selected from any of the 660 sequences fully defined
CC in the specification. A polynucleotide of the invention has cytostatic
CC activity, and may have a use in gene therapy, or in a vaccine. The
CC recombinant nucleic acids and polypeptides are useful for treating
CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
CC sarcomas. The present sequence represents a mouse mRNA of the invention.
XX
SQ Sequence 1729 BP; 466 A; 484 C; 436 G; 343 T; 0 U; 0 Other;

```

```

Query Match 39.4%; Score 570.2; DB 9; Length 1729;
Best Local Similarity 66.8%; Pred. No. 2.8e-176;
Matches 917; Conservative 0; Mismatches 423; Indels 33; Gaps 6;
QY 63 ATGGCAAGCATTATTAAAGGAGCATTTATCAGTGTGAGTGAAGACAGCAGTCTTGTGTGAG 122
Db 245 ATGGACGGACATTATTAAAGGAGCTCTGCTGTGTGAGTGAAGACAGTCTTGTGTGAG 304
QY 123 TGTGCTCAGGATGCG---CCCCACCTTGAAGAAGACAGAAATGACAGCTCTCTTCCAGT 179
Db 305 TCAGCATACGAGCGCGCAGCCCATCTCCCAAGGACAGATATGACTGCTTCGGGGAGTCT 364
QY 180 GAATATGGCAACATCAAGATGAGCCGCGCTTCCACAGAGCTGTTATACAG 239
Db 365 GACTACGGCAGCCCAAAATCAACCCCTGCGCACCGCAGAGTGTGATCAACAG 424
QY 240 CCCCCGCGCAGAGTTACCATTAAGATGGAGTGTAAACCAACACAGTTAATGGGTCAAG 299
Db 425 CC-----AGTGAGAGTCAATGTCAAGCGGAGTATGACCATGAATGGATCCAG 475
QY 300 AATTCACTGATGACTGCGAGCGTGCACAAAGAGGAGGAAATGTTAGCAGTTCAGACAAT 359
Db 476 GAGTCTCCGCTGAGTGCAGTGTCAAGCAAAATGTAACAGCTGCTGGGCGGAGCGGAGCC 535
QY 360 GTTGGGATGAACCTATGGAAGCTACATGGAAGAGAGCA---TATTCCGCTCCAAATATG 416
Db 536 AACCCCATGAACATAATAGCTACATGGATGAGAAGACGCGCCCTCTCTCCCAACATG 595
QY 417 ACACCAATGAACAGAGTATTGTGCCAGAGATCTTACGTTATGGAGCAGACACCAT 476
Db 596 ACCACCAAGACGGAGAGTCAATGTGCTGAGACCCCACTGTGGACAGAGAGCAC 655
QY 477 GTACGCGAGTGTGGAGTGGCAGTGAAGGAGTATGGTCTTCCAGAGCTGACATCTTG 536
Db 656 GTTGGACAGTGTGGAGTGGCTATAAGGAATACGGATTCATGGAGATTGACACTTCC 715
QY 537 TTGTTCCAGAACATTGATGGGAAAGTGTGTAATGACCAACCAAGATGACTTCCAGAG 596
Db 716 TTCTTCCAGAACATTGATGGGAAAGTGTGTAATGACCAACCAAGATGACTTCCAGAG 775
QY 597 CTCACGCCGAGCTATAACGACAGATATCTCTCTGTGCACACTACCTACCTCAGAGAGAG 656
Db 776 GCCACCTCCGCTACACACAGAGTGTGTTGTCGACCTCAGTTACCTCAGGGAAAG- 834
QY 657 GGAGCCACTTTTATTTTCCAAATACATCAGTTTACCCGAGAGCAACGCAAGAAATAACA 716
Db 835 -----TTCACTGCTGCGCTATAACCAACCTCCCATACAGACCACTCTCTCAGACTGAAT 889

```

QY 717 ACAAGGCCAGATTACCTTATGACCAAGCAGGAGATCAGCGTGGACGAGTCAAC---GC 773  
Db 890 GTCAGGAGAGACCCCTTCTTATGACTCTGTGAGGAGAGGAGCATGGAACAATAATATGAAC 949  
QY 774 CATCCACTCAGTCAAAAGCTACCAACCACTCATCTTCAAGTGCCTCAAAACAGAGAC 833  
Db 950 TCTGGCCTCAACAAGACCTCTCTCTGGAGATACAGACCATGGCAAGAACACTGAG 1009  
QY 834 CAGGCTCTCAGTTAGATCTTATCAGATTCTTGGACCGACGACCGCTCTTGCAGAAAT 893  
Db 1010 CAGCGGCCACGACGATCTTATCAGATCTGTGGGCCAACACGACGCGCTAGCAAAAC 1069  
QY 894 CCAGGAGTGGCAGATACACTATGGCAGTCTCTACTGAGCTTCTGTGGCAGCTCC 953  
Db 1070 CTTGGAGTGGCAGATCCAGCTGTGCGAGTTCTCTCTGGAACTACTGTCCGACAGCGCC 1129  
QY 954 AACTCCAACTGATCACCCTGGGAGGACAAATGGGAGTTCAAGATGACAGACCCCTGAT 1013  
Db 1130 AACGCCAGCTGTATCACCCTGGAGGGGACCAACGGGAGTTCAAAATGACGGACCTGAT 1189  
QY 1014 GAAGTGGTCCGGCTGGGAGAGAGAAAGCAAACTTAACATGAATGACAAACTC 1073  
Db 1190 GAGGTGGCCAGCGCTGGGAGAGCGGAAGCAAGCCCAACATGAATATGACAACTG 1249  
QY 1074 AGCGTGACATCTGCTACTATGACAAATAATATGACTAAAGTTCTATGGTAAACGC 1133  
Db 1250 AGCCGGCCCTCCGATCTACTATGACAAACAACTTATGACAAAGTGCATGGCAAGG 1309  
QY 1134 TATGCTCAAAATTTGATTTCCAGGAAATCGCTCAGGCGCTCCAGCTTACCTTCCAGAA 1193  
Db 1310 TATGCTCAAAATTTGATTTCCAGGAAATCGCTCAGGCGCTCCAGCTTACCTTCCAGAA 1369  
QY 1194 TCATCCATGACAAATCCCATCAGACCTCCCTTACATGAGTCTCTACCATGACACCCC 1253  
Db 1370 ACATCCATGACAAATCCCATCAGACCTCCCTTACATGAGTCTCTACCATGACACCCC 1429  
QY 1254 CAGAGATGAACTTTGTAGCTCCCATCCCTGCTTTGCCCGTAACTCATCCAGTTT 1313  
Db 1430 CAGAGAGTGAATTTGTGCTCCGCTTCACTCCATCTCCATGCTGCTGCTCCAGCTTC 1489  
QY 1314 TTTGCTGCCCTAATCCATCTGGAATTCACCACTGAGGAGCATCTACCCCAATACGAG 1373  
Db 1490 TTTGAGCAGCATCAATCAATCTGAGCTCCCGCACTGCTGGATCTATCCAAACCCAGT 1549  
QY 1374 CTGCCAGTGTCTATA-----TGCTTTCCCATCTTTGGCAGCTTACTACTA 1417  
Db 1550 GTCCCCCGCATCTTCAACCCACGCTGCTTACACTTAGGAGCTACTACTA 1602

RESULT 14  
AAQ50644  
ID AAQ50644 standard; cDNA; 2938 BP.  
XX AC AAQ50644;  
XX XX  
DT 25-MAR-2003 (revised)  
DT 26-MAY-1994 (first entry)  
XX DE Human Hum-Flu-1 gene clone BM025.  
XX DE  
XX KW chromosomal translocation; chimeric; chimeraic; Ewing sarcoma; Ews gene;  
KW malignant melanoma; hum-flu-1;  
KW primitive peripheral neuroectodermal tumour; human chromosome 11;  
KW human chromosome 22; ss.  
XX OS Homo sapiens.  
XX XX  
FH Key Location/Qualifiers  
FT CDS 143..1501  
FT FT /\*tag= a  
FT FT /product= "HUM-FLI-1"  
FT FT 2908..2913  
FT poly\_a\_signal /\*tag= b

XX WO9323549-A2.  
XX 25-NOV-1993.  
XX 19-MAY-1993; 93WO-FR000494.  
XX 20-MAY-1992; 92FR-00006123.  
XX (CNRS) CNRS CENT NAT RECH SCI.  
XX Aurias A, Delattre O, Desmaziere C, Melot T, Peter M, Plougastel B;  
XX Thomas G, Zucman J;  
XX WPI; 1993-386590/48.  
XX P-PSDB; AAR44556.  
XX New nucleic acid of EWS gene and its hybrid(s) - contg. gene sequence  
XX involved in chromosomal translocation, also derived mRNA, probes, fusion  
XX proteins etc., for diagnosis and treatment of Ewing sarcoma and melanoma.  
XX Disclosure; Fig 7; 123pp; French.  
XX The probe 11R1 was used to screen a human marrow cDNA library (Clontech  
XX cat.# HL1058). The clone BM025 was identified and sequenced. It  
XX represents the entire coding region together with 5' and 3'-UTRs of the  
XX Hum-Flu-1 gene. (Updated on 25-MAR-2003 to correct PN field.)  
XX Sequence 2938 BP; 847 A; 692 C; 675 G; 724 T; 0 U; 0 Other;  
Query Match 39.0%; Score 563.8; DB 2; Length 2938;  
Best Local Similarity 66.5%; Pred. No. 58-174;  
Matches 913; Conservative 0; Mismatches 427; Indels 33; Gaps 6;  
QY 63 ATGGCAAGCAGCTATTAAAGGAGCATTTATCAGTGGTGGAGTGAAGACAGTCTTGTGTGAG 122  
Db 143 ATGGAGCGGACTATTAAAGGAGGCTCTGTGCTGGTGGAGGAGCAGCAGCAGTCTTGTGAC 202  
QY 123 TGTGCTTACGATCG---CCCCACCTTGCAGAGACAGAGAAATGACAGCTCTTCCAGT 179  
Db 203 TCAGCTACGGAGCGCAGCCATCTCCCAAGCGCAGATGACTGCTCGGGAGTCTCT 262  
QY 180 GAATATGGGCAACATCAAGATGAGCCCGGTTCCCGCAGCAGGACTGTTATCACAG 239  
Db 263 GACTACGGGAGCCCAAGATCAACCCCTCCACAGCAGCAGGAGTGGATCAATCAG 322  
QY 240 CCCCCGCGCAGGTTTACCATTAAGATGGAGTGAACCCAAACCCAGCTTAAATGGTCAAGG 299  
Db 323 CC-----AGTGGGTCAACGTCAAGCGGAGTATGACCATGAATGATCCAGG 373  
QY 300 AATTCACTGATGACTGCGAGCGTGGCAAAAGGAGGAGAAATGGTTAGAGTTCAGACAT 359  
Db 374 GAGTCTCCGCTGACTGCGAGCGTTAGCAAAATGCGCAAGCTGGTGGCGGAGCGAGTCC 433  
QY 360 GTTGGGATGAATGAGAGTATGGAAGCTACATGGAAGAGAGAGCAT---ATTCCGCTCCAAATATG 416  
Db 434 AACCCATGAATCAACAGCTATATGGAGAGAGAAATGGCCCCCTCTCTCCACATG 493  
QY 417 ACAACCAATGAACGAGAGATTATGTGCCAGCAGATCTTACGTTATGGAGCAGACCAT 476  
Db 494 ACCACCAACGAGAGGAGAGTCACTCGTCCCGCAGAGCCCACTGTGGACACAGGAGCAT 553  
QY 477 GTAGGAGTGGCTGAGTGGGAGTGGAGTGGTCTTCCAGACGTTGGACATCTTG 536  
Db 554 GTAGGCAATGGCTGGAGTGGGCCATTAAGAGATATAGCTGTATGGAGATCGACATCC 613  
QY 537 TTGTTCCAGAACATTTGATGGGAAAGAGTTGTGTAAATGACAAAGATGATCTTCCAGAGA 596  
Db 614 TTTTCCAGAACATTTGATGGGAAAGAGTGTGTAAATGAAACAGGAGGAGTCTTCTCCGC 673  
QY 597 CTACGCGGAGCTATTAACGAGATATCTCTCTGTCACCTTACACTACCTCAGAGAGA 656  
Db 674 GCCACCAACCTCTACCAACAGGAGTGTCTTGTTCACCTTACCTTACCTCAGGGAAG- 732

QY 657 GGAGCCACTTTTATTTTCCAAATACATCAGTTTACCCGAGAAGCAACGCAAGAATAACA 716  
Db 733 -----TTCACTGCTGGCCATAAATACAACTCCACACCGACCAATCTCTCAGATTGAGT 787  
QY 717 ACAGGCCAGATTATACCTTATGAGCAGCAGGAGATCAGGTGGCAGAGTCACAGCCAT 776  
Db 788 GTCAAGAGAGACCCCTCTTATGACTCAGTCAGAGAGAGAGCAGTCGGGCAATACATGAAT 847  
QY 777 CC---CACTCAGTCAAAAGCTACCCAAACCATCATCTTCAACAGTGCCTCCAAACAGAGAC 833  
Db 848 TCTGGCCTCAACAAAGTCTCCCTTGGAGGGGCACAAACGATCAGTAAGATACAGAG 907  
QY 834 CAGGCTCTCAGTATGATCTTATCAGATTCTTGGACGGACAGCAGCCGCTCTTGGAAAT 893  
Db 908 CAACGGCCCCAGCAGATCCGATATCAGATCTGGGCGCGACACAGCAGTCGCTAGCCAAC 967  
QY 894 CCAGGAGTGGCGAGATACAGTATGGCAGTTCCTACTTGGAGCTTCTGTGGCAGAGTCC 953  
Db 968 CTTGGAAGCGGAGATCCAGCTGTGCAATCTCTCTGGAGCTGCTCTCCGACAGGCC 1027  
QY 954 AACTCCAACTGCATCCTCGGAGGCGCAAAATGGGAGTTCAAGATGACAGACCTTGAT 1013  
Db 1028 AACGCCAGCTGTATCACCTGGGAGGGACCAACGGGAGTTCAAATGACGGACCCCGAT 1087  
QY 1014 GAAGTGGCTCGGCTGGGAGAGAGGAAAGCAAACTACATGAATGACAAACTC 1073  
Db 1088 GAGTGGCCAGCGCTGGGGGAGCGGAAAGCAAGCCCAACATGAATTACGACAGCTG 1147  
QY 1074 AGCGTGCACTTCGCTACTATGACAAAAATATTATGACTAAAGTTCTATGTAACGC 1133  
Db 1148 AGCGGGCCCTCCGTTATTACTATGATAAAACATTATGACAAAGTGCACGGCAAAAGA 1207  
QY 1134 TATGCTCAAAATTTGATTTCCAGGATCCTCAGGCCCTCCAGCCTCCAGCTCCAGAA 1193  
Db 1208 TATGTTTCAAAATTTGATTTCCAGGCAATGCCAGGCTCTGCGAGCCATCCGACCGAG 1267  
QY 1194 TCATCCATGTACAAATACCATCAGACCTCCCTACATGAGTTCTTACCATGACACCCC 1253  
Db 1268 TCGTCCATGTACAGTACCTCTGACATCTCTTACATGCTCTTCTACCATGCTCCACG 1327  
QY 1254 CAGAAGATGAATTTGTAGTCTCCCATCCCTGCTTTGGCCGTAACCTCATCAGCTTT 1313  
Db 1328 CAGAAGTGAATTTGTCTCCCTCCCATCCCTCCATCCCTCCATGCTGTCACCTCCAGCTTC 1387  
QY 1314 TTTGCTGCCCTTAATCCATCTGGAATTCACCAACTGGAGGCATCTACCCCAATACCA-- 1371  
Db 1388 TTTGAGCGGATACATACATCTGACCTCCCGGAGGGAATCTACCCCAACCCCAAC 1447  
QY 1372 -----GGCTGCCAGTCTCATATGCTTCCCATCTTTGGCACTTACTACTACTA 1417  
Db 1448 GTCCCCCGCCATCCTTAACACCCACGTCGCTTACACTTTAGGAGCTACTACTA 1500

RESULT 15  
ID ADA02692 standard; cDNA; 1359 BP.  
XX  
AC ADA02692;  
XX  
DT 06-NOV-2003 (first entry)  
XX  
DE Human FLII carcinoma associated coding sequence, SEQ ID NO:1210.  
XX  
KW Human; carcinoma associated; oncogene; carcinoma; cancer; breast;  
KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;  
KW gene; ss.  
XX  
OS Homo sapiens.  
XX  
FN WO2003057146-A2.  
XX  
PD 17-JUL-2003.

XX 26-DEC-2002; 2002WO-US041414.  
PF 26-DEC-2001; 2001US-00035832.  
PR (SAGR-) SAGRES DISCOVERY.  
PA Morris DW;  
XX WPI; 2003-587068/55.  
DR  
XX  
XX New recombinant nucleic acid encoding carcinoma associated protein,  
PT useful for preparing compositions for treating carcinomas.  
XX  
XX Claim 1; SEQ ID NO 1210; 245pp; English.  
XX  
XX The invention relates to recombinant carcinoma associated (CA) nucleic  
CC acid sequences from mouse and human (ADA01482-ADA03094), and to  
CC recombinant carcinoma associated proteins (CAP) encoded by them. The  
CC invention also encompasses expression vectors and host cells comprising a  
CC CA nucleic acid, a polypeptide (especially an antibody) that specifically  
CC binds to the protein, and a biochip comprising CA nucleic acid or  
CC fragments thereof. The sequences of the invention were identified using  
CC oncogenic retroviruses, which insert into the genome of the host organism  
CC at random. Many of these do not carry transduced host oncogenes or  
CC pathogenic trans-acting viral genes, meaning that cancer incidence is a  
CC direct consequence of the effects of proviral integration into host  
CC protooncogenes. The CA nucleic acid sequences can be used to diagnose  
CC carcinoma (especially breast cancer, prostate cancer, lymphoma or  
CC leukaemia) or a propensity to carcinoma by determination of the sequence  
CC of a CA gene, or by determination of CA gene expression in particular  
CC tissues. CA nucleic acids, proteins and antibodies are also useful as  
CC therapeutic agents and in screening and evaluating drug candidates. The  
CC present sequence represents a specifically claimed human CA nucleic acid  
CC sequence of the invention. Note: The complete sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 1359 BP; 359 A; 422 C; 333 G; 245 T; 0 U; 0 Other;  
Query Match 38.9%; Score 562.2; DB 8; Length 1359;  
Best Local Similarity 66.4%; Pred. No. 16-173;  
Matches 912; Conservative 0; Mismatches 428; Indels 33; Gaps 6;  
QY 63 ATGGCAAGCATTATTAAAGAACATTATCAGTGGTGAAGTGAAGACAGTCTTGTGAG 122  
Db 1 ATGGACGGGACTATTAAAGGAGGCTCTGCTGGTGTGAGCGAGCAGTCCCTCTTTGAC 60  
QY 123 TGTGCTTACGGATCG---CCCCACCTTGCAAGACAGAAATGACAGCTCTCTTCCAGT 179  
Db 61 TCAGCGTACGGAGCGGCGAGCCCATCTCCCAAGGCGCAGATGACTGCTCGGGAGTCT 120  
QY 180 GAATATGGGCAAAACATCAAGATGAGCCGCGCTTCCCCAGCAGAGTCTGGTTATCACAG 239  
Db 121 GACTACGGGCGAGCCGCCACAGATCAACCCCTCCACACAGCAGAGTGGATCAATCAG 180  
QY 240 CCCCCGGCAGAGTTACCATTAAGATGAGTGAACCCAAACAGGTTAATGGTCAAGG 299  
Db 181 CC-----AGTGAGGGTCAACGTCGAAGCGGAGTATGACCAATGAATGATCCAGG 231  
QY 300 AATTCACCTGATGACTGACGCGTGGCAAAAGAGGAGGAAAATGTTAGCAGTTTCAACAAT 359  
Db 232 GAGTCTCCGGTGGACTGACGCGTTAGCAAAATGCAGCAAGCTGTGGCGGAGCGAGTCC 291  
QY 360 GTTGGGATGAATATGAGAGCTTACATGGAAGAGAGAGCAT---ATTCCGCTCCAAATATG 416  
Db 292 AACCCCATGAATACAAACAGCTATATGGACGAGAGAAATGGCCCCCTCTCTCCCAATG 351  
QY 417 ACAACCAATGAACGAAGAGTTATTGTGCCAGCAGATCTTACGTTATGGAGCAGACCAT 476  
Db 352 ACCACCAACGAGAGGAGAGTCTATCGTCCCGCAGAGCCCCACACTGTGGACAGAGGAT 411

QY 477 GTACGGCAGTGGCTGGAGTGGAGTGAAGAGTATGGTCTTCAGAGCGTGGACATCTTG 536  
 Db |||||  
 412 GTGAGGCAATGGCTGGAGTGGCCATAAAGGAGTACAGCTTGATGGAGATCGACACATCC 471  
 QY 537 TTGTTCCAGAAATATGATGGGAAGAGTGTGTAATAATGACCAAGATGACTTCCAGAGA 596  
 Db |||||  
 472 TTTTTCAGAAATATGATGGGAAGAGTGTGTAATAATGACCAAGAGTGTGTTCTCCCGC 531  
 QY 597 CTCACGGCAGCTATAACGACAGATATCTCTCTGTACACACCTACACTACCTCAGAGAGAGA 656  
 Db |||||  
 532 GCCACACCTCTACAACACGGAAGTGTGTGTACACACCTCAGTTACCTCAGGGAAG- 590  
 QY 657 GGAGCCATTTTATTTTCCAAATACATCAGTTTACCCAGAGCAACGCAAGAAATAACA 716  
 Db |||||  
 591 -----TTCACGTGCTGGCCCTATAATACACCTCCACACACCGACCAATCCTCACGATTGAGT 645  
 QY 717 ACAAGCCAGATTTTACCTTATGAGCAAGCAGGAGAGATCAGCGTGGACGAGTCAACAGCCAT 776  
 Db |||||  
 646 GTCAAAGAGAACCTTCTTATGACTCAGTCAAGAGAGAGCTTGGGGCAATTAACATGAAT 705  
 QY 777 CC---CACTCAGTCAAAAGCTACCCAAACCATCATCTTCAACAGTGGCCCAAAACAGAGAC 833  
 Db |||||  
 706 TCTGSCCTCAACAAAGTCTCCCTCTGGAGGGGCACAAACGATCAGTAAGNATACAGAG 765  
 QY 834 CAGGTTCTCAGTTAGATCTTATCAGATTTCTTGGACGGACCCAGCGCTCTTGCAAT 893  
 Db |||||  
 766 CAACGGCCCCAGCCAGATCCGTATCAGATCTCTGGGCCCCGACACGACGATCGCCTAGCCAAC 825  
 QY 894 CCAGGAGTGGCGAGATACAGTATGSCAGTTCTTCTTGGAGCTTCTGTGGACAGCTCC 953  
 Db |||||  
 826 CCTGGACGCGGAGATCCAGCTGTGGCAATCTCTCTGGAGCTGCTCTCCGACAGCGCC 885  
 QY 954 AACTCCAACTGCATCACTGGGAGGGCACAAATGGGAGTTCAAGATGACAGACCTTGAT 1013  
 Db |||||  
 886 AACGCCAGCTATCACCTGGGAGGGACCAACGGGGAGTTCAAAATGACGGACCCCGAT 945  
 QY 1014 GAAGTGGCTCGCGTGGGAGAGAGGAAAGCAAACTTACATGAATGACAAATC 1073  
 Db |||||  
 946 GAGTGGCCAGCGCTGGGCGAGCGGAAAGCAAGCCCAACATGAATTTACGACAGCTG 1005  
 QY 1074 AGCGTGCACTTCGCTACTACTATGACAAATAATTTATGACTAAAGTTGATGTPAAACGC 1133  
 Db |||||  
 1006 AGCGGGCCCTCCGTTATTACTATGATAAAACATTTATGACCAAGTGCACGGCAAAAGA 1065  
 QY 1134 TATGCTACAAATTTGATTTCCAGGATCGCTCAGGCGCTCCAGCGCTCACCTCCAGAA 1193  
 Db |||||  
 1066 TATGCTTACAAATTTGATTTCCAGGATTCGCCAGCTTCGCCAGCTTCGACGCCACATCCGACCGAG 1125  
 QY 1194 TCATCCATGTACAAATACCCATCAGACCTCCCTTACATGAGTTCCCTACCATGCAACCCC 1253  
 Db |||||  
 1126 TCGTCCATGTACAGTACCTTCTGACATCTCTACATGCTTCTTACCATGCCATGCCACCGAG 1185  
 QY 1254 CAGAGATGAATTTGTAGCTCCCATCCCTGCTTGGCCGTAACTCATCCAGCTTT 1313  
 Db |||||  
 1186 CAGAGGTGAATTTGTCTCTCCCATCCCATCTCCATGCTGCTCTCTCTCCAGCTTC 1245  
 QY 1314 TTTGTGCCCCCTAATCCATATCGGAATTCACCAACTGGAGGCACTTACCCCAATACCA-- 1371  
 Db |||||  
 1246 TTTGGAGCCGATCACATATCTGGACCTCCCGACGGGGGAATCTACCCCAACCCCAAC 1305  
 QY 1372 -----GGCTGCCAGCTGCTCATATGCTTCCCATCTTGGCACCTACTACTA 1417  
 Db |||||  
 1306 GTCCCCGCCCATCTTAACACCCAGTGGCTTTCACACTTAGGCAGCTACTACTA 1358

Search completed: May 27, 2004, 16:39:39  
 Job time : 687 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 27, 2004, 16:28:07 ; Search time 143 Seconds  
(without alignments)  
5615.482 Million cell updates/sec

Title: US-09-902-772-1  
Perfect score: 1447  
Sequence: 1 gaattccggaacgaataa.....gaaagaagcgcaagaaaa 1447

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 27747546 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/6C\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/6D\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1447	100.0	1447	3	US-08-878-177-1
2	1356	93.7	1528	3	US-08-878-177-3
3	563.8	39.0	2938	2	US-08-343-443B-3
4	180	12.4	1752	3	US-09-360-779-1
5	180	12.4	1752	4	US-09-435-335-1
6	148.4	10.3	2532	4	US-08-620-312D-869
7	124.6	8.6	1604	5	US-08-306-691B-43
8	124.6	8.6	1604	5	PCT-US93-06251-9
9	122.2	8.4	2268	3	US-09-344-579-1
10	117.4	8.1	2667	2	US-08-469-412A-1
11	117.4	8.1	2667	3	US-09-021-715-1
12	115.2	8.0	328	2	US-08-343-443B-5
13	100.6	7.0	1933	4	US-09-920-759-3
14	100.6	7.0	1976	4	US-09-920-759-10
15	96	6.6	2064	3	US-08-875-944B-1
16	96	6.6	2064	3	US-09-116-049-3
17	96	6.6	2064	4	US-03-602-868A-1
18	96	6.6	2064	4	US-09-884-363-3
19	95.2	6.6	2410	2	US-08-780-835B-1
20	95.2	6.6	2410	3	US-09-303-268-1
21	95.2	6.6	2410	3	US-09-116-049-1
22	95.2	6.6	2410	4	US-09-884-363-1
23	95.2	6.6	2410	4	US-09-328-925-13
24	93.4	6.5	665	2	US-09-920-759-11
25	88.6	6.1	2266	2	US-09-213-767-1
26	87.4	6.0	2544	2	US-08-469-412A-6
27	87.4	6.0	2544	3	US-09-021-715-6

## ALIGNMENTS

### RESULT 1

US-08-878-177-1  
; Sequence 1, Application US/08878177  
; Patent No. 6294354  
; GENERAL INFORMATION:  
; APPLICANT: Chugai Seiyaku Kabushiki Kaisha, 5001, Iwamoto et  
; TITLE OF INVENTION: Cell Calcification Suppressing Proteins and Genes of  
; FILE REFERENCE: chugai seiyaku kabushiki kaisha 5001  
; CURRENT APPLICATION NUMBER: US/08/878,177  
; CURRENT FILING DATE: 1997-06-18  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 1447  
; TYPE: DNA  
; ORGANISM: C-11 gene, c-erg gene w/ deletion, chicken DNA  
US-08-878-177-1

Query Match	100.0%;	Score 1447;	DB 3;	Length 1447;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1447;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	GAATTCGCGAAGCAATAATATTATTAGCAATATTAGCGATCAATATCTTGATCACA	60	Sequence 13, Appli
Db	1	GAATTCGCGAAGCAATAATATTATTAGCAATATTAGCGATCAATATCTTGATCACA	60	Sequence 3, Appli
QY	61	TTATGGCAAGCACTATTAAAGGAGCATTTATCATGTGGTGAAGACCAAGTCTTGTGTTG	120	Sequence 44, Appli
Db	61	TTATGGCAAGCACTATTAAAGGAGCATTTATCATGTGGTGAAGACCAAGTCTTGTGTTG	120	Sequence 44, Appli
QY	121	AGTGTGCTTACGGATCGCCCACTTGCAGACAGAAATGACAGCTCTCTTCCAGTG	180	Sequence 44, Appli
Db	121	AGTGTGCTTACGGATCGCCCACTTGCAGACAGAAATGACAGCTCTCTTCCAGTG	180	Sequence 44, Appli
QY	181	AATATGGCAAGCACTATTAAAGGAGCATTTATCATGTGGTGAAGACCAAGTCTTGTGTTG	240	Sequence 44, Appli
Db	181	AATATGGCAAGCACTATTAAAGGAGCATTTATCATGTGGTGAAGACCAAGTCTTGTGTTG	240	Sequence 44, Appli
QY	241	CCCCGCGCAGAGTTACCAATTAAGATGGAGTGAACCCAAACAGGTTAATGGGTCAAGCA	300	Sequence 44, Appli
Db	241	CCCCGCGCAGAGTTACCAATTAAGATGGAGTGAACCCAAACAGGTTAATGGGTCAAGCA	300	Sequence 44, Appli
QY	301	ATTACCTGATGACTGCAGCGTGGCAAAAGAGGAGGAAATGTTAGAGTTTACAGCAATG	360	Sequence 44, Appli
Db	301	ATTACCTGATGACTGCAGCGTGGCAAAAGAGGAGGAAATGTTAGAGTTTACAGCAATG	360	Sequence 44, Appli
QY	361	TTGGGATGAACATGAGAGCTACATGGAGAGCAATATTCGGCTCCAAATATGACAA	420	Sequence 44, Appli
Db	361	TTGGGATGAACATGAGAGCTACATGGAGAGCAATATTCGGCTCCAAATATGACAA	420	Sequence 44, Appli



```
QY 421 CCAATGAACGAAGAGTTATTGTCGACGAGATCTAGCTTATGGAGCAGACGACATGTAC 480
Db 421 CCAATGAACGAAGAGTTATTGTCGACGAGATCTAGCTTATGGAGCAGACGACATGTAC 480
QY 481 GGCAGTGGCTGGAGTGGGCACTGAAGAGTATGTCCTCCAGAGCTGGACATCTTGTGT 540
Db 481 GGCAGTGGCTGGAGTGGGCACTGAAGAGTATGTCCTCCAGAGCTGGACATCTTGTGT 540
QY 541 TCCAGAACATTTGATGGGAAGAGTTGTGTAATATGACCAAGATGACTTCCAGAGACTCA 600
Db 541 TCCAGAACATTTGATGGGAAGAGTTGTGTAATATGACCAAGATGACTTCCAGAGACTCA 600
QY 601 GCGCGAGCTATAACGACATATCTCTGTGCACACCTACACTACTCTCAGAGAGAGGAG 660
Db 601 GCGCGAGCTATAACGACATATCTCTGTGCACACCTACACTACTCTCAGAGAGAGGAG 660
QY 661 CCACCTTTTATTTTCCAAATACATCATGTTTACCAGAAAGCAACGCAAGAAATACACAA 720
Db 661 CCACCTTTTATTTTCCAAATACATCATGTTTACCAGAAAGCAACGCAAGAAATACACAA 720
QY 721 GGCAGATTTTACCTTATGAGCAACGAGGAGATCAGCGTGCAGAGTCACAGCATCCCA 780
Db 721 GGCAGATTTTACCTTATGAGCAACGAGGAGATCAGCGTGCAGAGTCACAGCATCCCA 780
QY 781 CTCAGTCAAAAGCTACCCAACTCATCTTCAACAGTGGCCCAAAACAGAGACGAGGTC 840
Db 781 CTCAGTCAAAAGCTACCCAACTCATCTTCAACAGTGGCCCAAAACAGAGACGAGGTC 840
QY 841 CTCAGTTAGATCCCTTATCAGATCTTGGACCGACGACGCGCTTGGCAAAATCCAGGGA 900
Db 841 CTCAGTTAGATCCCTTATCAGATCTTGGACCGACGACGCGCTTGGCAAAATCCAGGGA 900
QY 901 GTGGGCGATACAGCTATGGCAGTCTCTTCTGAGAGTCTCTGCGACAGCTCCAACTCCA 960
Db 901 GTGGGCGATACAGCTATGGCAGTCTCTTCTGAGAGTCTCTGCGACAGCTCCAACTCCA 960
QY 961 ACTGCATCAGCTGGAGGAGCAAAATGGGAGTTCAAGATGACAGACCTCTGATGAGTGG 1020
Db 961 ACTGCATCAGCTGGAGGAGCAAAATGGGAGTTCAAGATGACAGACCTCTGATGAGTGG 1020
QY 1021 CTCGGGCTGGGAGAGAGGAAAGCAAACTAACTGAATGACAACTGACAACTCAGCGGTG 1080
Db 1021 CTCGGGCTGGGAGAGAGGAAAGCAAACTAACTGAATGACAACTGACAACTCAGCGGTG 1080
QY 1081 CACTTCGCTACTACTATGACAAAATATTATGACTAAAGTTTCTGTTAAACGCTATGCTT 1140
Db 1081 CACTTCGCTACTACTATGACAAAATATTATGACTAAAGTTTCTGTTAAACGCTATGCTT 1140
QY 1141 ACAAAATTTGATTTCCACGGATCGCTCAGGCGCTCCAGCGCTCCAGCTCCAGAACTATCCA 1200
Db 1141 ACAAAATTTGATTTCCACGGATCGCTCAGGCGCTCCAGCGCTCCAGCTCCAGAACTATCCA 1200
QY 1201 TGTTCAAAATACCAATCAGACCTCCCTTACATGAGTCTCTTACATGCAACCCCGAGAGA 1260
Db 1201 TGTTCAAAATACCAATCAGACCTCCCTTACATGAGTCTCTTACATGCAACCCCGAGAGA 1260
QY 1261 TGAATCTTTGAGTCCCTATCCCTGCTTTGCGCGCTAACTCTCATCCAGCTTTTGTGCTG 1320
Db 1261 TGAATCTTTGAGTCCCTATCCCTGCTTTGCGCGCTAACTCTCATCCAGCTTTTGTGCTG 1320
QY 1321 CCCCTAATCCATCTGGAATTTCAACAACTGAGGCACTTACCCCAATACAGGCTGCCAG 1380
Db 1321 CCCCTAATCCATCTGGAATTTCAACAACTGAGGCACTTACCCCAATACAGGCTGCCAG 1380
QY 1381 CTGCTCATCTGCTCCCTATCTTGGCACTTACTACTAGTGGGAAAGAAAGAGCGCC 1440
Db 1381 CTGCTCATCTGCTCCCTATCTTGGCACTTACTACTAGTGGGAAAGAAAGAGCGCC 1440
QY 1441 AAGAAA 1447
Db 1441 AAGAAA 1447
```

```
RESULT 2
US-08-878-177-3
; Sequence 3, Application US/08878177
; Patent No. 6294354
; GENERAL INFORMATION:
; APPLICANT: Chugai Seiyaku Kabushiki Kaisha, 5001, Iwamoto et
; TITLE OF INVENTION: Cell Calcification Suppressing Proteins and Genes of
; TITLE OF INVENTION: the Proteins
; FILE REFERENCE: chugai seiyaku kabushiki kaisha 5001
; CURRENT APPLICATION NUMBER: US/08/878,177
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1528
; TYPE: DNA
; ORGANISM: c-erg gene, chicken DNA
; US-08-878-177-3

Query Match      93.7%; Score 1356; DB 3; Length 1528;
Best Local Similarity 94.7%; Pred. No. 0;
Matches 1447; Conservative 0; Mismatches 0; Indels 81; Gaps 1;

QY 1 GAATTCGCGAAGCAATTAATTATTAGCAATTTATTAGCGATCAATTAATCTTGATCACA 60
Db 1 GAATTCGCGAAGCAATTAATTATTAGCAATTTATTAGCGATCAATTAATCTTGATCACA 60
QY 61 TTATGGCAAGCACTATTAAAGGAAGCATTATCAGTGGTGAAGCAAGCAGTCTTGTGTTG 120
Db 61 TTATGGCAAGCACTATTAAAGGAAGCATTATCAGTGGTGAAGCAAGCAGTCTTGTGTTG 120
QY 121 AGTGTGCTACGGATCGCCCGACCTTGCAGAGAGAAATGACAGCTCTCTTCCAGTG 180
Db 121 AGTGTGCTACGGATCGCCCGACCTTGCAGAGAGAAATGACAGCTCTCTTCCAGTG 180
QY 181 AATATGGGCAAAACATCAAAAGATGAGCCGCGCTTCCCGAGCAGACTGGTTATCACAGC 240
Db 181 AATATGGGCAAAACATCAAAAGATGAGCCGCGCTTCCCGAGCAGACTGGTTATCACAGC 240
QY 241 CCCCGCCAGAGTTACCAATTAAGATGGAGTGAACCCAAAACAGTTAATGGGTCAAGGA 300
Db 241 CCCCGCCAGAGTTACCAATTAAGATGGAGTGAACCCAAAACAGTTAATGGGTCAAGGA 300
QY 301 ATTCACTTGATGACATGCAAGCTGCGCAAAAGAGGAGAAATAGTTAGCAAGTTTCAACAATG 360
Db 301 ATTCACTTGATGACATGCAAGCTGCGCAAAAGAGGAGAAATAGTTAGCAAGTTTCAACAATG 360
QY 361 TTGGGATGAATCTATGGAAGCTACATGGAAGAGAGCATATTCGCGCTCCAAATATGACAA 420
Db 361 TTGGGATGAATCTATGGAAGCTACATGGAAGAGAGCATATTCGCGCTCCAAATATGACAA 420
QY 421 CCAATGCAAGAGAGTTATTGTCGAGCAGATCTTACGTTATGGAGCAGACAGACCATGTAC 480
Db 421 CCAATGCAAGAGAGTTATTGTCGAGCAGATCTTACGTTATGGAGCAGACAGACCATGTAC 480
QY 481 GGCAGTGGCTGGAGTGGGCACTGAAGAGTATGTCCTTCCAGAGCTGGACATCTTGTGT 540
Db 481 GGCAGTGGCTGGAGTGGGCACTGAAGAGTATGTCCTTCCAGAGCTGGACATCTTGTGT 540
QY 541 TCCAGAACATTTGATGGGAAGAGTTGTGTAATATGACCAAGATGACTTCCAGAGACTCA 600
Db 541 TCCAGAACATTTGATGGGAAGAGTTGTGTAATATGACCAAGATGACTTCCAGAGACTCA 600
QY 601 GCGCGAGCTATAACGACATATCTCTGTGCACACCTACACTACTCTCAGAGAGAGGAG 654
Db 601 GCGCGAGCTATAACGACATATCTCTGTGCACACCTACACTACTCTCAGAGAGAGGAG 654
QY 655 -----
Db 655 -----
QY 661 TTCCACATTTGACTTCCAGATGATGTGATAAGCGCTTACAAACCTCTCCACGGTTAATGC 720
Db 661 TTCCACATTTGACTTCCAGATGATGTGATAAGCGCTTACAAACCTCTCCACGGTTAATGC 720
QY 655 -----GAGGAGCGCACTTTTATTTTCCAAATACATCATCAGTTTACCCAGAG 699
```





Query Match 12.4%; Score 180; DB 4; Length 1752;  
Best Local Similarity 74.0%; Pred. No. 3.2e-51;  
Matches 228; Conservative 0; Mismatches 80; Indels 0; Gaps 0;  
QY 879 AGCGTCTTGGCAATCCAGGAGTGGGAGATACAGCTATGGGAGTTCCTACTGAGCTT 938  
Db 529 AGCCCTGGGTACAGAAAGGAGCGGGAGATCCAGTTGTGGCAGTTTCTACTGAGCTG 598  
QY 939 CTGTGGGACAGCTCCAACTCCAACTGCATCACCTGGGAGGAGCAAAATGGGAGTTCAAG 998  
Db 599 CTGGGAGACCGCGGAACCGCGGTGCATCGCGTGGGAGGCGGCCACGCGAGTTCAAG 648  
QY 999 ATGACAGACCTGTATGAAGTGGCTCGGCTGGGAGGAGAGAAAACCAACCTTAACATG 1058  
Db 649 CTCACGACCCGACGAGTGGCGGAGCTGGGCGGAGCGACGAGGCCCAATATG 708  
QY 1059 AACTATGACAACTCAGCGCTGCTACTGCTACTATGATGACAAAATATATGACTAAA 1118  
Db 709 AACTACGACAAAGTAAAGTCGAGCACTGGCTACTACTACGACAAAACATCATGAGCAAG 768  
QY 1119 GTTCATGTAAACGTATGCCCTACAAATTTGATTTCACCGGAATCGCTCAGGCCCTCCAG 1178  
Db 769 GTGCACGCAAGCGCTACGCTACCGCTTTCACCTTCAGGCGCTGGCACAGGCTTGCAG 828  
QY 1179 CCTCACCC 1186  
Db 829 CCACCACC 836

RESULT 6  
US-09-620-312D-869  
; Sequence 869, Application US/09620312D  
; Patent No. 6569662  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yungqing  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: John Tillinghast  
; APPLICANT: Drmanac, Radjoe T.  
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
; FILE REFERENCE: 784CIP2B  
; CURRENT APPLICATION NUMBER: US/09/620,312D  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1105  
; SOFTWARE: dt\_FL\_genes Version 1.0  
; SEQ ID NO 869  
; LENGTH: 2532  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(2532)  
US-09-620-312D-869

Query Match 10.3%; Score 148.4; DB 4; Length 2532;

Best Local Similarity 73.5%; Pred. No. 3.4e-40;  
Matches 216; Conservative 0; Mismatches 76; Indels 2; Gaps 2;  
QY 891 AATCCAGGAGTGGGAGATACAGCTATGGGAGTTCCTACTGAGCTTCTGTCGACAGC 950  
Db 2091 AATGATGACAGGAGCAGATCCAGCTGTGGCAGTTTCTGCTGGAGCTGCTGCTGACCCG 2150  
QY 951 TCAAACTCCAACTGCATCACCTGGGAGGAGCAAAATGGGAGTTCAGAGATGACAGACCT 1010  
Db 2151 GCGACGCGG-CTGATCGGTGGGAGGCGGTCCGCGGAGTTCAAGCTCAGGACCG 2209  
QY 1011 GATGAAGTGGCTGGGCTGGGAGGAGAGAAAACCAACCTAAATGAATGATGACAAA 1070  
Db 2210 GACGAGTGGCGCGGTGGGAGGAGCGGCAAGAGCAAGCCCAACATGAATACGACAA 2269  
QY 1071 CTCAGCGGTGCACTTGGCTACTACTATGACAAAATATATGACTAAAGTTTCATGTAAA 1130  
Db 2270 CTGAGCGGCGCTGTAT-CTACTACGACAAACATCATGAGCAAGTGCATGTTAAG 2328  
QY 1131 CGCTATGCTACAAATTTGATTTCACCGGAATCGCTCAGGCCCTCCAGCCTCAC 1184  
Db 2329 CGTACGCTACCGCTTCGACTTCCAGGCGCTGGCGAGGCGGCGGCTCTAC 2382

RESULT 7  
US-08-306-691B-43  
; Sequence 43, Application US/08306691B  
; Patent No. 5734039  
; GENERAL INFORMATION:  
; APPLICANT: Galabretta, Bruno  
; APPLICANT: Skorski, Tomasz  
; TITLE OF INVENTION: ANTISENSE  
; TITLE OF INVENTION: OLIGONUCLEOTIDES TARGETING COOPERATING ONCOGENES  
; NUMBER OF SEQUENCES: 55  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Seidel, Gonda, Lavorigna & Monaco, P.C.  
; STREET: Two Penn Center, Suite 1800  
; CITY: Philadelphia  
; STATE: Pennsylvania  
; COUNTRY: U.S.A.  
; ZIP: 19102  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WordPerfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/306,691B  
; FILING DATE: September 15, 1994  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Monaco, Daniel A.  
; REGISTRATION NUMBER: 30,480  
; REFERENCE/DOCKET NUMBER: 8321-8  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (215) 568-8383  
; TELEFAX: (215) 568-5549  
; TELEX: NO. 5734039e  
; INFORMATION FOR SEQ ID NO: 43:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1604 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-306-691B-43

Query Match 8.6%; Score 124.6; DB 1; Length 1604;  
Best Local Similarity 61.6%; Pred. No. 4.1e-32;  
Matches 199; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 835 AGCGTCTCAGTAGATCCTTATCAGATTTCTTGACCGACAGCGCGCTTTGCAATC 894  
 Db 1207 ACCGTGTGACCTCAATAAGGACAAGCGCTGTCATTCTGCTGCGCTAGCTGGGTACA 1266  
 QY 895 CAGGAGTGGCAGATACAGCTATGCGAGTTCCTACTGAGCTTCTGTGGCAGCTCCA 954  
 Db 1267 CAGGAGTGGCAGATACAGCTATGCGAGTTCCTACTGAGCTTCTGTGGCAGCTCCA 1326  
 QY 955 ACTCCAACTGCATCACCTCTGGAGGGCAAAATGGGGAGTTTCAAGATGACAGACCTCTGATG 1014  
 Db 1327 GTGAGTCTTTTATCAGCTGGACAGAGATGGCTGGGAATTCAAACTTTCTGACCCAGATG 1386  
 QY 1015 AAGTGGCTCGCGTGGGAGAGAGAAAGCAACCTACATGACACTATGACAACTCA 1074  
 Db 1387 AGGTGGCCAGAGATGGGAAAGAGGAAACAAACCTTAAGATCAATTTATGAGAACTGA 1446  
 QY 1075 GCCGTGCACTTCGCTACTACTATGACAAATAATTTATGACTAAAGTTTCATGGTAAACGCT 1134  
 Db 1447 GCCGTGCGCTACGCTACTATTACGACAAATAATTTATGACTAAAGTTTCATGGTAAACGCT 1506  
 QY 1135 ATGCTACAAATTTGATTTCCAC 1157  
 Db 1507 ACGTGTACCGCTTTGTGTGAC 1529

RESULT 8  
 PCT-US93-06251-9  
 ; Sequence 9, Application PC/TUS9306251  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wickstrom, Eric and Rife, Jason P.  
 ; TITLE OF INVENTION: Trivalent Synthesis of Oligonucleotides Containing  
 ; ; TITLE OF INVENTION: Stereospecific Alkylphosphonates and Arylphosphonates  
 ; NUMBER OF SEQUENCES: 93  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER  
 ; STREET: 400 Garden City Plaza  
 ; CITY: Garden City  
 ; STATE: NY  
 ; COUNTRY: USA  
 ; ZIP: 11530  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US93/06251  
 ; FILING DATE: 19930630  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Digiglio, Frank S.  
 ; REGISTRATION NUMBER: 31,346  
 ; REFERENCE/DOCKET NUMBER: 8586  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 516-742-4343  
 ; TELEFAX: 516-742-4366  
 ; TELEX: 230 901 SANS UR  
 ; INFORMATION FOR SEQ ID NO: 9:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1604 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; PCT-US93-06251-9

Query Match 8.6%; Score 124.6; DB 5; Length 1604;  
 Best Local Similarity 61.6%; Pred. No. 4.1e-32;  
 Matches 199; Conservative 0; Mismatches 124; Indels 0; Gaps 0;  
 QY 835 ACCGTCTCAGTAGATCCTTATCAGATTTCTTGACCGACAGCGCGCTTTGCAATC 894  
 Db 1207 ACCGTGTGACCTCAATAAGGACAAGCGCTGTCATTCTGCTGCGCTAGCTGGGTACA 1266

QY 895 CAGGAGTGGCAGATACAGCTATGCGAGTTCCTACTGAGCTTCTGTGGCAGCTCCA 954  
 Db 1267 CAGGAGTGGCAGATACAGCTATGCGAGTTCCTACTGAGCTTCTGTGGCAGCTCCA 1326  
 QY 955 ACTCCAACTGCATCACCTCTGGAGGGCAAAATGGGGAGTTTCAAGATGACAGACCTCTGATG 1014  
 Db 1327 GTGAGTCTTTTATCAGCTGGACAGAGATGGCTGGGAATTCAAACTTTCTGACCCAGATG 1386  
 QY 1015 AAGTGGCTCGCGTGGGAGAGAGAAAGCAACCTACATGACACTATGACAACTCA 1074  
 Db 1387 AGGTGGCCAGAGATGGGAAAGAGGAAACAAACCTTAAGATCAATTTATGAGAACTGA 1446  
 QY 1075 GCCGTGCACTTCGCTACTACTATGACAAATAATTTATGACTAAAGTTTCATGGTAAACGCT 1134  
 Db 1447 GCCGTGCGCTACGCTACTATTACGACAAATAATTTATGACTAAAGTTTCATGGTAAACGCT 1506  
 QY 1135 ATGCTACAAATTTGATTTCCAC 1157  
 Db 1507 ACGTGTACCGCTTTGTGTGAC 1529

RESULT 9  
 US-09-344-579-1  
 ; Sequence 1, Application US/09344579  
 ; Patent No. 6054316  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Brenda F. Baker  
 ; APPLICANT: Lex M. Cowsett  
 ; TITLE OF INVENTION: ANTISENSE MODULATION OF ETS-2 EXPRESSION  
 ; FILE REFERENCE: RTS-0063  
 ; CURRENT APPLICATION NUMBER: US/09/344,579  
 ; CURRENT FILING DATE: 1999-06-25  
 ; NUMBER OF SEQ ID NOS: 47  
 ; SEQ ID NO 1  
 ; LENGTH: 2268  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (291)..(1700)  
 ; US-09-344-579-1

Query Match 8.4%; Score 122.2; DB 3; Length 2268;  
 Best Local Similarity 66.5%; Pred. No. 3.6e-31;  
 Matches 175; Conservative 0; Mismatches 88; Indels 0; Gaps 0;  
 QY 895 CAGGAGTGGCAGATACAGCTATGCGAGTTCCTACTGAGCTTCTGTGGCAGCTCCA 954  
 Db 1363 CAGGAGTGGCAGCTATTTCAGCTGTGGCAGTTTCTCTGGAGCTGCTATCAGACAAATCCT 1422  
 QY 955 ACTCCAACTGCATCACCTCTGGAGGGCAAAATGGGGAGTTTCAAGATGACAGACCTCTGATG 1014  
 Db 1423 GCCAGTCAATTCATCAGCTGGACTGGAGAGCGGATGGAGTTTAAAGCTCGCCGCCCGGATG 1482  
 QY 1015 AAGTGGCTCGCGTGGGAGAGAGAAAGCAACCTAAACATGAATGAACTATGACAACTCA 1074  
 Db 1483 AGGTGGCCCGCGTGGGAAAGAGGAAATAATGACCCAGATGAACTACGAGAGCTGA 1542  
 QY 1075 GCCGTGCACTTCGCTACTACTATGACAAATAATTTATGACTAAAGTTTCATGGTAAACGCT 1134  
 Db 1543 GCCGGGCTTACGCTACTATTACGACAAATAATTTATGACTAAAGTTTCATGGTAAACGCT 1602  
 QY 1135 ATGCTACAAATTTGATTTCCAC 1157  
 Db 1603 ACGTGTACCGCTTTGTGTGAC 1625

RESULT 10  
 US-08-469-412A-1  
 ; Sequence 1, Application US/08469412A  
 ; Patent No. 5856125  
 ; GENERAL INFORMATION:

```
; APPLICANT: Mavrothalassitis, George J.
; APPLICANT: Blair, Donald G.
; APPLICANT: Fisher, Robert J.
; APPLICANT: Beal Jr., Gregory J.
; APPLICANT: Athanasiou, Meropi A.
; APPLICANT: Sgouras, Dionysios N.
; TITLE OF INVENTION: The ERF Genetic Locus and Its Products
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS: 16
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,412A
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Garrett-Wackowski, Eugenia
; REGISTRATION NUMBER: 37,330
; REFERENCE/DOCKET NUMBER: 015280-229000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2667 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 123..1769
; OTHER INFORMATION: /note= "human ERF (ETS2 Repressor
; OTHER INFORMATION: Factor) cDNA"
US-08-469-412A-1

Query Match      8.1%; Score 117.4; DB 2; Length 2667;
Best Local Similarity 63.1%; Pred. No. 1.8e-29;
Matches 181; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 894 CCAGGAGTGGCAGATACAGCTATGGCAGTTCTCTGAGCTTCTGTGGACAGCTCC 953
DB 186 CTGGCTCAAGGCAGATCCAGCTGTGGCACTTTATCTCTGGAGCTGCTGCGAAGGAG 245

QY 954 AACTCCAACTGCATCACCTGGGAGGCAAAATGGGAGTTTCAAGATGACAGACCTTGAT 1013
DB 246 TACAGGGCGTTCATTCCTGCGAGGGGACTACGGGGAATTCGTCAACAGACCTTGAT 305

QY 1014 GAAGTGGCTCGCGGTGGGAGAGAGAAAGCAAACTTAACATGAACTATGACAACTC 1073
DB 306 GAGGTGGCCGGCTGTGGGGGTTTCGCAAGTGCAGGCCCGAGATGAATTAGCACAAGCTG 365

QY 1074 AGCGGTGCATTCGCTACTACTATACAAAAATATTATGACTAAAGTTTCATGGTAAACGC 1133
DB 366 AGCGGGCCCTCGGCTATTACTATACAGCGCAATTCGACACAGACCAAGGGAACGG 425

QY 1134 TATGCTCAAAATTTGATTTTCCACGGAATCGCTACGCCCTCCAGCC 1180
DB 426 TTCACTCAAGTTCAATTTCAACAAACTGGTGTGTCATTAATACCC 472

RESULT 11
US-09-021-715-1
; Sequence 1, Application US/09021715
```

```
; Patent No. 6194547
; GENERAL INFORMATION:
; APPLICANT: Mavrothalassitis, George J.
; APPLICANT: Blair, Donald G.
; APPLICANT: Fisher, Robert J.
; APPLICANT: Beal Jr., Gregory J.
; APPLICANT: Athanasiou, Meropi A.
; APPLICANT: Sgouras, Dionysios N.
; TITLE OF INVENTION: The ERF Genetic Locus and Its Products
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS: 16
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/021,715
; FILING DATE: 10-Feb-1998
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Garrett-Wackowski, Eugenia
; REGISTRATION NUMBER: 37,330
; REFERENCE/DOCKET NUMBER: 015280-229000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2667 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 123..1769
; OTHER INFORMATION: /note= "human ERF (ETS2 Repressor
; OTHER INFORMATION: Factor) cDNA"
US-09-021-715-1

Query Match      8.1%; Score 117.4; DB 3; Length 2667;
Best Local Similarity 63.1%; Pred. No. 1.8e-29;
Matches 181; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 894 CCAGGAGTGGCAGATACAGCTATGGCAGTTCTCTGAGCTTCTGTGGACAGCTCC 953
DB 186 CTGGCTCAAGGCAGATCCAGCTGTGGCACTTTATCTCTGGAGCTGCTGCGAAGGAG 245

QY 954 AACTCCAACTGCATCACCTGGGAGGCAAAATGGGAGTTTCAAGATGACAGACCTTGAT 1013
DB 246 TACAGGGCGTTCATTCCTGCGAGGGGACTACGGGGAATTCGTCAACAGACCTTGAT 305

QY 1014 GAAGTGGCTCGCGGTGGGAGAGAGAAAGCAAACTTAACATGAACTATGACAACTC 1073
DB 306 GAGGTGGCCGGCTGTGGGGGTTTCGCAAGTGCAGGCCCGAGATGAATTAGCACAAGCTG 365

QY 1074 AGCGGTGCATTCGCTACTACTATGACAAAAATATTATGACTAAAGTTTCATGGTAAACGC 1133
DB 366 AGCGGGCCCTCGGCTATTACTATACAGCGCAATTCGACACAGACCAAGGGAACGG 425

QY 1134 TATGCTCAAAATTTGATTTTCCACGGAATCGCTACGCCCTCCAGCC 1180
DB 426 TTCACTCAAGTTCAATTTCAACAAACTGGTGTGTCATTAATACCC 472
```

RESULT 12  
 US-08-343-443B-5  
 ; Sequence 5, Application US/08343443B  
 ; Patent No. 5968734  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Aurias, Alain  
 ; APPLICANT: Delattre, Olivier  
 ; APPLICANT: Desmaze, Chantal  
 ; APPLICANT: Melot, Thomas  
 ; APPLICANT: Peter, Martine  
 ; APPLICANT: Ploougaatel, Beatrice  
 ; APPLICANT: Thomas, Gilles  
 ; APPLICANT: Zucman, Jessica  
 ; TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF  
 ; CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL  
 ; TRANSLOCATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS  
 ; TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID  
 ; TRANSLOCATIONS  
 ; NUMBER OF SEQUENCES: 129  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Weiser & Associates  
 ; STREET: 230 South Fifteenth Street  
 ; CITY: Philadelphia  
 ; STATE: PA  
 ; COUNTRY: USA  
 ; ZIP: 19102  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: AEDIT 1.0 DOS text editor  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/343.443B  
 ; FILING DATE: 18-NOV-1994  
 ; CLASSIFICATION: 514  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/FR93/00494  
 ; FILING DATE: 19-MAY-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: FR 92/06123  
 ; FILING DATE: 20-MAY-1992  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Weiser, Gerard J.  
 ; REGISTRATION NUMBER: 19,763  
 ; REFERENCE/DOCKET NUMBER: 989.6121P  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 215-875-8383  
 ; TELEFAX: 215-875-8394  
 ; INFORMATION FOR SEQ ID NO: 5:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 328 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 1..327  
 ; US-08-343-443B-5  
 Query Match 8.0%; Score 115.2; DB 2; Length 328;  
 Best Local Similarity 67.4%; Pred. No. 2.3e-29;  
 Matches 178; Conservative 0; Mismatches 83; Indels 3; Gaps 1;  
 QY 733 CTTATGACCAACGAGGAGATCAGCGTGGACGAGTCACAGCCATCC---CACTCAGTCAA 789  
 DB 65 CTTATGACTCAGTCAGAGAGAGAGTTGGGCAATAACATGAATCTGGCCCTCAACAAA 124  
 QY 790 AAGCTACCCCAACCATCTCTTCAACAGTGGCCCAACAGACAGCGTCTCTCAGTTAG 849  
 DB 125 GTCTCTCCCTTGGAGGGGCACAAACGATCAGTAAATACAGAGCAACGGCCCGCAG 184  
 QY 850 ATCTCTATCAGATTTCTTGACCGACACAGCGCTTGTGCAAAATCCAGGGAGTGGGAGA 909

Db 185 ATCCGTATCAGATCTCTGGGCCGACGACGATCGCTAGCCAAACCTTGAAGGGGCGAGA 244  
 QY 910 TACAGCTATGGCAGTTCTTACTGGAGCTTCTGTGGACAGTCCAACTCCAACTGCATCA 969  
 Db 245 TCAGCTGTGGCAATTCCTCTGGAGCTGCTCTCCGACAGCGCCAAACGCCAGCTGTATCA 304  
 QY 970 CTTGGGAGGGCACAATATGGGAGT 993  
 Db 305 CTTGGGAGGGGACCAACAGGGGAGT 328  
 RESULT 13  
 US-09-920-759-3  
 ; Sequence 3, Application US/09920759  
 ; Patent No. 6537811  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Brenda F. Baker  
 ; APPLICANT: Susan M. Freier  
 ; TITLE OF INVENTION: ANTISENSE MODULATION OF SAP-1 EXPRESSION  
 ; FILE REFERENCE: RTS-0267  
 ; CURRENT APPLICATION NUMBER: US/09/920,759  
 ; CURRENT FILING DATE: 2001-08-01  
 ; NUMBER OF SEQ ID NOS: 91  
 ; SEQ ID NO 3  
 ; LENGTH: 1933  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (150)...(1445)  
 ; US-09-920-759-3  
 Query Match 7.0%; Score 100.6; DB 4; Length 1933;  
 Best Local Similarity 64.3%; Pred. No. 9.2e-24;  
 Matches 151; Conservative 0; Mismatches 84; Indels 0; Gaps 0;  
 QY 915 CTATGGCAGTTCTTACTGGAGCTTCTGTGGACAGCTCCAACTCCAACTGCATCACCTGG 974  
 Db 168 CTGTGGCAGTTCTTCTTTCAGCTCTCTGAGAGCCCTCAGAACAGACATGATCTGTGG 227  
 QY 975 GAGGGCACAATGGGGAGTTCAAGATGACAGACCCCTGATGAAGTGGCTCGCGTTGGGA 1034  
 Db 228 ACCTCTAATGATGGGCAGTTTAAAGCTTTTTCAGGAGAGAGTGGCTCTCTGGGGG 287  
 QY 1035 GAGAGAAAAGCAACCTTACATGAACTATGACAACTCAGCGGTGCATTCGCTACTAC 1094  
 Db 288 ATTCCGACAGACAGCCCTAACATGAATTATGACAACTCAGCGGAGCCCTCAGATCAT 347  
 QY 1095 TATGACAAAATATTATGACTAAAGTTTCATGTTAAACGCTATGCTTACAAATTTG 1149  
 Db 348 TATGTAAGCATATCATCAAAAAGTGAATGTCAGAAAGTTTGTGTACAAGTTG 402  
 RESULT 14  
 US-09-920-759-10  
 ; Sequence 10, Application US/09920759  
 ; Patent No. 6537811  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Brenda F. Baker  
 ; APPLICANT: Susan M. Freier  
 ; TITLE OF INVENTION: ANTISENSE MODULATION OF SAP-1 EXPRESSION  
 ; FILE REFERENCE: RTS-0267  
 ; CURRENT APPLICATION NUMBER: US/09/920,759  
 ; CURRENT FILING DATE: 2001-08-01  
 ; NUMBER OF SEQ ID NOS: 91  
 ; SEQ ID NO 10  
 ; LENGTH: 1976  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (150)...(1367)  
 ; US-09-920-759-10

Query Match 7.0%; Score 100.6; DB 4; Length 1976;  
Best Local Similarity 64.3%; Pred. No. 9.4e-24;  
Matches 151; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 915 CTATGGCAGTTCTCTACTGGAGTTCTGTGCGACAGCTCCAACTGCACTGCTGG 974  
DB 168 CTGGGCGAGTTCTCTCTGAGAGCTCTGAGAGCTCAGAAAGCAGCATGATCTGTGG 227  
QY 975 GAGGCGACAAATGGGAGTTCAAGATGACAGACCTGATGAAGTGGCTGCGCGTTGGGGA 1034  
DB 228 ACCTCTAATGATGGGCGAGTTTAAAGCTTTTGAGGCGAGAGAGTGGCTCTCTGCGGG 287  
QY 1035 GAGAGGAAAGCAACCTTAACATGACTATGACAACTCAGCCCTGCGACTTCGCTACTAC 1094  
DB 288 ATTCCAGAGCAAGCCCTAAATGATGATTTATGACAACTCAGCCGAGCCCTCAGATATAT 347  
QY 1095 TATGACAAAATATTTATGACTTAAAGTTTCATGCTTAAAGCTATGCTTACAAAATTTG 1149  
DB 348 TATGTAAGAATATCATCAAAAAGTGAATGGTCAGAAGTTTGTGTACAAAGTTTG 402

## RESULT 15

US-08-875-944B-1  
; Sequence 1, Application US/08875944B  
; Patent No. 6096542  
; GENERAL INFORMATION:  
; APPLICANT: FUJINAGA, Kei  
; APPLICANT: YOSHIDA, Koichi  
; APPLICANT: HIGASHINO, Fumihiro  
; TITLE OF INVENTION: CANCER CONTROL  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: BROWDY AND NEIMARK, P.L.L.C.  
; STREET: 624 Ninth Street N.W., Ste. 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20001  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/875,944B  
; FILING DATE: 07-AUG-1997  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 07-020173  
; FILING DATE: 08-FEB-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/JP96/00016  
; FILING DATE: 09-JAN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BROWDY, Roger L.  
; REGISTRATION NUMBER: 25,618  
; REFERENCE/DOCKET NUMBER: FUJINAGA-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 628-5197  
; TELEFAX: (202) 737-3528  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2084 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1386

US-08-875-944B-1

Query Match 6.6%; Score 96; DB 3; Length 2064;  
Best Local Similarity 62.5%; Pred. No. 3.7e-22;  
Matches 150; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 910 TACAGCTATGGCAGTTCTCTACTGAGCTTCTGTGCGACAGCTCCAACTGCAATCA 969  
DB 956 TGCAGCTGTGGCAATTTCTGTGGCTTCTGTGATGACCAACAATGCCATTTCAATTG 1015  
QY 970 CCTGGAGGGCACAATGGGAGTTTCAAGATGACAGACCTGATGAAGTGGCTCGCGGTT 1029  
DB 1016 CCTGGACGGGCGGGGAATGGAGTTTCAAGCTCATTTGAGCTGAGGAGTGGCAGGCTCT 1075  
QY 1030 GGGGAGAGAGGAAAGCAAACTTAACATGAATATGACAAACTCAGCCCTGCACTTCGCT 1089  
DB 1076 GGGGCACTCCAGAGAACCGGCCAGCCATGATTTACGACAGCTGAGCCGCTCGCTCCGAT 1135  
QY 1090 ACTACTATGACAAAATATTTATGACTTAAAGTTTCATGCTTAAAGCTATGCTTAAATTTG 1149  
DB 1136 ACTATTATGAGAAAGGCATCATGACAGAGTGGCTGGTGGCGTTTACGTGTACAAGTTTG 1195

Search completed: May 27, 2004, 19:27:30  
Job time : 147 secs





QY 181 AATATGGGCAACATCAAGATGAGCGCGCGTTCGCCAGCAGGACTGGTTATCAAGC 240  
Db 181 AATATGGGCAACATCAAGATGAGCGCGCGTTCGCCAGCAGGACTGGTTATCAAGC 240  
QY 241 CCCCGCCAGAGTTACCAATTAAGATGAGTGTAAACCAACAGGTTAATGGGTCAAGGA 300  
Db 241 CCCCGCCAGAGTTACCAATTAAGATGAGTGTAAACCAACAGGTTAATGGGTCAAGGA 300  
QY 301 ATTCAACCTGATGACTGACGCGTGGCAAAAGAGGAGGAAATGGTTAGCAGTTCAGACAATG 360  
Db 301 ATTCAACCTGATGACTGACGCGTGGCAAAAGAGGAGGAAATGGTTAGCAGTTCAGACAATG 360  
QY 361 TTGGATGAACATATGGAAGCTACATGGAAGAGAGCATATTCGCGCTCCAAATATGACAA 420  
Db 361 TTGGATGAACATATGGAAGCTACATGGAAGAGAGCATATTCGCGCTCCAAATATGACAA 420  
QY 421 CCAATGAACGAGAGTTATTTGTCGACGAGATCTAGCTTATGAGCAGACACCATGTAC 480  
Db 421 CCAATGAACGAGAGTTATTTGTCGACGAGATCTAGCTTATGAGCAGACACCATGTAC 480  
QY 481 GGCAGTGGCTGGAGTGGGCAAGGAGTATGCTTCCAGAGCTGGACATCTTGTGT 540  
Db 481 GGCAGTGGCTGGAGTGGGCAAGGAGTATGCTTCCAGAGCTGGACATCTTGTGT 540  
QY 541 TCCGAACATTTGATGGGAAGAGTTGTGTAATGACCAAGATGACTTCCAGAGCTCA 600  
Db 541 TCCGAACATTTGATGGGAAGAGTTGTGTAATGACCAAGATGACTTCCAGAGCTCA 600  
QY 601 CGCCAGAGCTAAGCAGATATCTCTGTGCACACCTACACTACCTCAGAGAGAGGAG 660  
Db 601 CGCCAGAGCTAAGCAGATATCTCTGTGCACACCTACACTACCTCAGAGAGAGGAG 660  
QY 661 CCACCTTTATTTTCCAAATACATCAGTTTACCAGAGCAACGCAAGATATACACAA 720  
Db 661 CCACCTTTATTTTCCAAATACATCAGTTTACCAGAGCAACGCAAGATATACACAA 720  
QY 721 GGCAGATTTTACCTTATGAGCAAGCGAGAGATCAGCGTGGACGAGTCACAGCCATCCCA 780  
Db 721 GGCAGATTTTACCTTATGAGCAAGCGAGAGATCAGCGTGGACGAGTCACAGCCATCCCA 780  
QY 781 CTCAGTCAAAAGCTACCAACATCTCTCAACAGTGGCCCAAAACAGAGACAGAGTC 840  
Db 781 CTCAGTCAAAAGCTACCAACATCTCTCAACAGTGGCCCAAAACAGAGACAGAGTC 840  
QY 841 CTCAGTATAGCTCTTATCAGATTTCTGGACCGACAGCAGCGCTCTTGCAATCCAGGA 900  
Db 841 CTCAGTATAGCTCTTATCAGATTTCTGGACCGACAGCAGCGCTCTTGCAATCCAGGA 900  
QY 901 GTGGCGAGATACAGTATGGCAGTTCTTCTGAGAGCTTCTGCGGACAGCTCCAACTCCA 960  
Db 901 GTGGCGAGATACAGTATGGCAGTTCTTCTGAGAGCTTCTGCGGACAGCTCCAACTCCA 960  
QY 961 ACTGCATCACCTGGGAGGSCAAATGGGAGTTCAAGATGACAGACCTGATGAAGTGG 1020  
Db 961 ACTGCATCACCTGGGAGGSCAAATGGGAGTTCAAGATGACAGACCTGATGAAGTGG 1020  
QY 1021 CTCGCGTGTGGGAGAGGAGAAAGCAACCTAACATGAATATGACAACTCAGCCGTG 1080  
Db 1021 CTCGCGTGTGGGAGAGGAGAAAGCAACCTAACATGAATATGACAACTCAGCCGTG 1080  
QY 1081 CACTTCGCTACTACTATGACAAAATATATGACTTAAGTTTCAATGAAGCTATGCCT 1140  
Db 1081 CACTTCGCTACTACTATGACAAAATATATGACTTAAGTTTCAATGAAGCTATGCCT 1140  
QY 1141 ACAAAATTTGATTTCCACGGAATCGCTCAGGCGCTTCCAGCCCTCAGCAATCATCCA 1200  
Db 1141 ACAAAATTTGATTTCCACGGAATCGCTCAGGCGCTTCCAGCCCTCAGCAATCATCCA 1200  
QY 1201 TGTACAAATACCATCAGACCTCCCTACATGAGTTCTTACATGACACACCCCAAGA 1260  
Db 1201 TGTACAAATACCATCAGACCTCCCTACATGAGTTCTTACATGACACACCCCAAGA 1260  
QY 1261 TGAACCTTTGTAGTCTCCCATCCCTGCTTTGCCCGTAACTCATCCAGCTTTTGTG 1320

Db 1261 TGAACCTTTGTAGTCTCCCATCCCTGCTTTGCCCGTAACTCATCCAGCTTTTGTG 1320  
QY 1321 CCCTTAATCCATACATGGAATTCACCACTGGAGGATCTACCCCAATACCAGCTGCCAG 1380  
Db 1321 CCCTTAATCCATACATGGAATTCACCACTGGAGGATCTACCCCAATACCAGCTGCCAG 1380  
QY 1381 CTGCTCATATGCTTCCCATCTTGGCAGCTACTACTAAGTGGGAAAGAAAGAAAGCGCC 1440  
Db 1381 CTGCTCATATGCTTCCCATCTTGGCAGCTACTACTAAGTGGGAAAGAAAGAAAGCGCC 1440  
QY 1441 AAGAAAA 1447  
Db 1441 AAGAAAA 1447  
  
RESULT 2  
US-09-902-772-3  
; Sequence 3, Application US/09902772  
; Patent No. US20020164739A1  
; GENERAL INFORMATION:  
; APPLICANT: Chugai Seiyaku Kabushiki Kaisha, 5001, Iwamoto et  
; TITLE OF INVENTION: Cell Calcification Suppressing Proteins and Genes of  
; TITLE OF INVENTION: the Proteins  
; FILE REFERENCE: chugai seiyaku kabushiki kaisha 5001  
; CURRENT APPLICATION NUMBER: US/09/902,772  
; CURRENT FILING DATE: 2001-07-12  
; PRIOR APPLICATION NUMBER: US/08/878,177  
; PRIOR FILING DATE: 1997-06-18  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 1528  
; TYPE: DNA  
; ORGANISM: c-ery gene, chicken DNA  
US-09-902-772-3  
  
Query Match 93.7%; Score 1356; DB 9; Length 1528;  
Best Local Similarity 94.7%; Pred. No. 0;  
Matches 1447; Conservative 0; Mismatches 0; Indels 81; Gaps 1;  
  
QY 1 GAATTCGGGAACGAATAATATTATTAGCAATATTAGCGATCAATATCTTGATCA 60  
Db 1 GAATTCGGGAACGAATAATATTATTAGCAATATTAGCGATCAATATCTTGATCA 60  
QY 61 TTATGCAAGCACTATTAAAGGAAGCATTTATCAGTGGTGAAGCAAGCAGTCTTGTG 120  
Db 61 TTATGCAAGCACTATTAAAGGAAGCATTTATCAGTGGTGAAGCAAGCAGTCTTGTG 120  
QY 121 AGTGTGCTTACGGATCGCCACCTTGCAAGACAGAAATGACAGCTCTCTTCCAGTG 180  
Db 121 AGTGTGCTTACGGATCGCCACCTTGCAAGACAGAAATGACAGCTCTCTTCCAGTG 180  
QY 181 AATATGGGCAACATCAAGATGAGCGCGGTTCCCGCAGGAGCTGGTTATCACAGC 240  
Db 181 AATATGGGCAACATCAAGATGAGCGCGGTTCCCGCAGGAGCTGGTTATCACAGC 240  
QY 241 CCCCGCCAGAGTTACCAATTAAGATGAGTGTAAACCAACAGGTTAATGGGTCAAGGA 300  
Db 241 CCCCGCCAGAGTTACCAATTAAGATGAGTGTAAACCAACAGGTTAATGGGTCAAGGA 300  
QY 301 ATTCACTCATGACTGACGCGTGGCAAAAGAGGAGGAAATGGTTAGCAGTTCAGACAATG 360  
Db 301 ATTCACTCATGACTGACGCGTGGCAAAAGAGGAGGAAATGGTTAGCAGTTCAGACAATG 360  
QY 361 TTGGATGAACATATGGAAGCTACATGGAAGAGAGCATATTCGCGCTCCAAATATGACAA 420  
Db 361 TTGGATGAACATATGGAAGCTACATGGAAGAGAGCATATTCGCGCTCCAAATATGACAA 420  
QY 421 CCAATGAACGAGAGTTATTTGTCGACGAGATCTAGCTTATGAGCAGACACCATGTAC 480  
Db 421 CCAATGAACGAGAGTTATTTGTCGACGAGATCTAGCTTATGAGCAGACACCATGTAC 480

481 GGCGAGTGGCTGGAGTGGGAGTGGAGTGGTCTTCCAGAGCTGGACATCTGTGT 540  
481 GGCGAGTGGCTGGAGTGGGAGTGGAGTGGTCTTCCAGAGCTGGACATCTGTGT 540  
541 TCCAGAACATTGATGGGAGAGTGGTAAATGACCAAGATGATTCAGAGACTCA 600  
541 TCCAGAACATTGATGGGAGAGTGGTAAATGACCAAGATGATTCAGAGACTCA 600  
601 GGCGAGTGTAAAGCAGATATCTCTGTGACACCTACACTACCTCAGAGAGA- 654  
601 GGCGAGTGTAAAGCAGATATCTCTGTGACACCTACACTACCTCAGAGAGACTCCT 660  
655 ----- 654  
661 TTCACATTGACTTTCAGATGATGTGATAAGGCTTACAAAATCTCCACGGTTAATGC 720  
655 -----GAGGAGCCACTTTTATTTTCCAAATACATCAGTTTACCCAGAAG 699  
721 ATGCTAGAAACACAGGAGAGCCACTTTTATTTTCCAAATACATCAGTTTACCCAGAAG 780  
700 CAACGCAAGAAATACAAAGGCGAGATTTACCTTATGACCAAGCGAGGATCAGCGT 759  
781 CAACGCAAGAAATACAAAGGCGAGATTTACCTTATGACCAAGCGAGGATCAGCGT 840  
760 GGCGAGTGCAGGCTACCTCAGTCAAAAGCTACCCACCATCTCTCAACAGTGC 819  
841 GGCGAGTGCAGGCTACCTCAGTCAAAAGCTACCCACCATCTCTCAACAGTGC 900  
820 CCAAAACAGAAAGCAGGCTCCTCAGTTAGATCTTATCAGATCTTTGGACCGACACGCA 879  
901 CCAAAACAGAAAGCAGGCTCCTCAGTTAGATCTTATCAGATCTTTGGACCGACACGCA 960  
880 GCGGTCTTGCAAATCCAGGAGTGGGAGATGAGTATGCGAGTTCTTACTGAGCTTC 939  
961 GCGGTCTTGCAAATCCAGGAGTGGGAGATGAGTATGCGAGTTCTTACTGAGCTTC 1020  
940 TGTGGGAGCTCCAACTCCAACTGATCCTGGGAGGACAAATGGGAGTTCAAGA 999  
1021 TGTGGGAGCTCCAACTCCAACTGATCCTGGGAGGACAAATGGGAGTTCAAGA 1080  
1000 TGACAGACCTTGATGAGTGGCTGGGCTTGGGAGAGAGAAAGCAAACTTAACTGA 1059  
1081 TGACAGACCTTGATGAGTGGCTGGGCTTGGGAGAGAGAAAGCAAACTTAACTGA 1140  
1060 ACTATGACAACTCAGCGCTGCTTCTGCTACTACTATGACAAATATTTATGACTAAG 1119  
1141 ACTATGACAACTCAGCGTGGCTTCTGCTACTACTATGACAAATATTTATGACTAAG 1200  
1120 TTCATGTTAAAGCGCTATCCCTACAAATTTGATTTCCAGGAAATCGCTCAGGCTCCAGC 1179  
1201 TTCATGTTAAAGCGCTATCCCTACAAATTTGATTTCCAGGAAATCGCTCAGGCTCCAGC 1260  
1180 CTCACCTCCAGATCATCTGATCAAAATACCCATCAGCTCCCTCAGATGATTCCT 1239  
1261 CTCACCTCCAGATCATCTGATCAAAATACCCATCAGCTCCCTCAGATGATTCCT 1320  
1240 ACCATGACACCCCGAGAGATGAACTTTGTAGCTCCCTCCTCCCTGCTTTGCGCGTAA 1299  
1321 ACCATGACACCCCGAGAGATGAACTTTGTAGCTCCCTCCTCCCTGCTTTGCGCGTAA 1380  
1300 COTCATCCAGCTTTTGTGCTCCCTTAATCCATACCTGGAATTCACCACTGGAGGATCT 1359  
1381 COTCATCCAGCTTTTGTGCTCCCTTAATCCATACCTGGAATTCACCACTGGAGGATCT 1440  
1360 ACCCAATACAGGCTGCGAGTCTCATATGCTTCCCTTCCCTTGGACCTTACTTAAG 1419  
1441 ACCCAATACAGGCTGCGAGTCTCATATGCTTCCCTTCCCTTGGACCTTACTTAAG 1500  
1420 TGGGGAAGAAAGAGCGCCCAAGAAA 1447  
1501 TGGGGAAGAAAGAGCGCCCAAGAAA 1528

RESULT 3  
US-10-211-462-41  
; Sequence 41, Application US/10211462  
; Publication No. US20040033495A1  
; GENERAL INFORMATION:  
; APPLICANT: Murray, Richard  
; APPLICANT: Glynn, Richard  
; APPLICANT: Watson, Susan R.  
; APPLICANT: Aziz, Natasha  
; APPLICANT: Ros Biotechnology, Inc.  
; TITLE OF INVENTION: Methods of Diagnosis of Angiogenesis, Compositions and  
; TITLE OF INVENTION: Methods of Screening for Angiogenesis Modulators  
; FILE REFERENCE: 018501-006200US  
; CURRENT APPLICATION NUMBER: US/10/211,462  
; CURRENT FILING DATE: 2003-02-13  
; PRIOR APPLICATION NUMBER: US 09/784,356  
; PRIOR FILING DATE: 2001-02-14  
; PRIOR APPLICATION NUMBER: US 09/791,390  
; PRIOR FILING DATE: 2001-02-22  
; PRIOR APPLICATION NUMBER: US 60/310,025  
; PRIOR FILING DATE: 2001-08-03  
; PRIOR APPLICATION NUMBER: US 60/334,244  
; PRIOR FILING DATE: 2001-11-29  
; NUMBER OF SEQ ID NOS: 230  
; SOFTWARE: Patent In Ver. 2.1.  
; SEQ ID NO 41  
; LENGTH: 1389  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-211-462-41  
Query Match 63.6%; Score 921; DB 13; Length 1389;  
Best Local Similarity 81.1%; Pred. No. 1.4e-294; Indels 12; Gaps 2;  
Matches 1100; Conservative 0; Mismatches 245;  
481 TATTAAGAGAGATTCATCAGTGGTGGAGTGGAGTGGTCTTCCAGTGAATATATGGGCAAA 193  
481 TATTAAGAGAGATTCATCAGTGGTGGAGTGGAGTGGTCTTCCAGTGAATATATGGGCAAA 193  
481 ATGCCCCACCTTGCAGAGAGATTCAGTGGTGGAGTGGTCTTCCAGTGAATATATGGGCAAA 193  
481 AAGCCACACCTTGGTGGTGGAGTGGAGTGGTCTTCCAGTGAATATATGGGCAAA 152  
481 ATCAAGAGATGAGCCCCGCGGTCTCCAGAGAGTGGTCTTCCAGTGAATATATGGGCAAA 253  
481 TTCCAGAGATGAGCCCCGCGGTCTCCAGAGAGTGGTCTTCCAGTGAATATATGGGCAAA 212  
481 TACCATTAAGTGGAGTGTAAACCAACAGGTTAATGGTCAAGGAATTCACCTGATGA 313  
481 CACCATCAAAATGGAAATGTAAACCTAGCCAGTGGTGAATGGTCAAGGAATTCACCTGATGA 272  
481 CTCAGCGTGGCAAAAGAGGAGGAAATGGTGGAGTGGTGGTGGAGTGAACCTA 373  
481 ATGCACTGTGGCAAAAGGCGGAGAGTGGTGGAGTGGTGGTGGAGTGAACCTA 332  
481 TGAAGCTCATGGAAGAGAGAGATTCGGCTCCAAATATGACCAATGGAAGAG 433  
481 CGGAGCTCATGGAAGAGAGAGATTCGGCTCCAAATATGACCAATGGAAGAG 392  
481 AGTTATGTGGCAGAGATTCCTACCTTATGGAGCAGACAGACCATGTACGGCAGTGGCTGGA 493  
481 AGTTATGTGGCAGAGATTCCTACCTTATGGAGCAGACAGACCATGTACGGCAGTGGCTGGA 452  
481 GTGGGCGGTGAAAGATATGGGCTTCCAGAGCTCAACATCTTGTATTCCAGAAACATCGA 512  
481 TGGGAAGAGTGTGTAAATGACCAAGATGACTTCCAGAGCTCAACATCTTGTATTCCAGAAACATCGA 613  
481 TGGGAAGAGTGTGTAAATGACCAAGATGACTTCCAGAGCTCAACATCTTGTATTCCAGAAACATCGA 572  
481 CGCAGATATCTCTCTCTCTCAGCTTCAACCTACACTACTCTCAGAGAGA-----GAGGAGCCAC 664

Db 573 GCGGACATCCTTCTCTCACATCTCCACTACCTCAGAGAGACTCCTCTTCCACATTTGAC 632  
Qy 665 TTTTATTTTCCAAATACATCAGTTTACCCAGAGAGCAACGACAGAGATTAACAAGGCC 724  
Db 633 TTCAGATGATGTTGATAAGCTTACAAAACCTCCACGGTTATGATGCTGAAACAC 692  
Qy 725 AGATTACCTTATGAGCAAGCGAGAGATCAGCGTGAAGAGTCCACAGCCATCCCA--C 781  
Db 693 AGATTACCATATGAGCCCCCAGAGATCAGCTGAGCCGGTCCAGGCCACCCACGCC 752  
Qy 782 TCAGTCAAAAGCTACCCCAACCATCATCTTCAACAGTGCACAAAACAGAAAGACAGCGTCC 841  
Db 753 CCAGTCGAAAGCTGCTCAACCATCTCTCTTCCAGAGTGCCTCAAACTGAAAGACAGCGTCC 812  
Qy 842 TCAGTTAGATCCTTATCAGATCTTGGACCGACGAGCGCGTCTTCAAAATCCAGGAG 901  
Db 813 TCAGTTAGATCCTTATCAGATCTTGGACCAACAGTAGCGCGCTTCAAAATCCAGGAG 872  
Qy 902 TGGGAGATACAGCTATGCGAGTCTTCTTCTGAGCTTCTGCGACAGCTCCCACTCCAA 961  
Db 873 TGGCCAGATCAGCTTGGCGAGTCTCTCTGAGAGTCTCTGCGACAGCTCCCACTCCAG 932  
Qy 962 CTGCATCACCTGGGAGGACAAATGGGAGTTCAGAGTACAGACCTGATGAAGTGGC 1021  
Db 933 CTGCATCACCTGGGAGGACAAACAGCGGAGTTCAGAGTACAGCTCCCGACGAGTGGC 992  
Qy 1022 TGGGCTTGGGAGAGAGAAAGCAACCTAACTGAACTATGAACTACAACTCAGCCGTGC 1081  
Db 993 CCGGCGCTGGGAGAGCGGAAGAGCAACCAACATGAATACGATGAAGTTCAGCGCGC 1052  
Qy 1082 ACTTCGCTACTATGACAAAAATATTATGACTAAAGTTCATGGTAAAGCTATGCTTA 1141  
Db 1053 CCTCGTTACTATGACAAAGACATCATGACCAAGTTCATGGGAAGGCTACGCTTA 1112  
Qy 1142 CAATTTGATTTCCAGGAATCGCTCAGGCCCTCAGGCTCACCTCCAGATCATCCAT 1201  
Db 1113 CAAGTTCAGTTTCCAGGGATCGGCCAGGCTCCAGGCCACCCCGCGAGTCACTCT 1172  
Qy 1202 GTACAAATACCATCAGACCTCCCTACATGAGTTCTTACCATGACACACCCCAAGAGAT 1261  
Db 1173 GTACAAATACCATCAGACCTCCCTACATGAGTTCTTACCATGACACACCCCAAGAGAT 1232  
Qy 1262 GAATTTGAGTCCCATCCCTCCCTGCTTTCGCGTAACTCATCTCAGCTTTTGTGTC 1321  
Db 1233 GAATTTGAGTCCCATCCCTCCCTGCTTTCGCGTAACTCATCTCAGCTTTTGTGTC 1292  
Qy 1322 CCCTAATCCATCGAATTCACCAACTGAGGAGCTTACCCCAATACCAAGTCCGAGC 1381  
Db 1293 CCCTAATCCATCGAATTCACCAACTGAGGAGCTTACCCCAATACCAAGTCCGAGC 1352  
Qy 1382 TGCTCATATGCTTCCCATCTTGGACCTACTACTAA 1418  
Db 1353 CAGCCATATGCTTCTCATCTGGGACCTACTACTAA 1389

## RESULT 4

US-10-087-192-1127  
; Sequence 1127, Application US/10087192  
; Publication No. US20020182586A1  
; GENERAL INFORMATION:  
; APPLICANT: Morris, David W.  
; APPLICANT: Engelhard, Eric K.  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
; FILE OF INVENTION: CANCER  
; FILE REFERENCE: 52945200122  
; CURRENT APPLICATION NUMBER: US/10/087,192  
; PRIORITY FILING DATE: 2002-03-01  
; PRIORITY FILING DATE: US 09/747,377  
; PRIORITY FILING DATE: 2000-12-22  
; PRIORITY FILING DATE: US 09/798,586  
; PRIORITY FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 2059  
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1127  
; LENGTH: 1389  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-087-192-1127

Query Match 63.6%; Score 921; DB 13; Length 1389;  
Best Local Similarity 81.1%; Pred. No. 1.4e-294;  
Matches 1100; Conservative 0; Mismatches 245; Indels 12; Gaps 2;

Qy 74 TATTAAAGGAAGCATTTATCAGTGTGAGTGAAGACAGTCTTGTGTTGAGTGTGCTACGG 133  
Db 33 TATCAGGAGAGCTTATCAGTTGTGAGTGAAGACAGTCTTGTGTTGAGTGTGCTACGG 92  
Qy 134 ATGCGCCCACTTGCAGAAAGACAGAAATGACAGCTCTCTTCCAGTGAATATGGGCAAC 193  
Db 93 AACGCCACACCTGGCTTAAGACAGAGATGACCGCTCTCTCCAGCGACTATGGACAGAC 152  
Qy 194 ATCAAGATGAGCCCGCGTTCACAGAGACAGTCTTGTGTTGAGTGTGCTACGG 253  
Db 153 TTCCAGATGAGCCCGCGTTCACAGAGATGAGTCTTGTGTTGAGTGTGCTACGG 212  
Qy 254 TACCATTAAAGTGGAGTGTAAACCAACAGGTTTAAATGGTCAAGGAATTCACCTGATGA 313  
Db 213 CACCATCAAAATGGAATGTAACTTAGCCAGGTGAATGGCTCAAGGAATCTCTCTGATGA 272  
Qy 314 CTGCAGCGTGGCAAAAGAGGAGAAATGGTTAGCAGTTCAGACCAATGTTGGGATGACTA 373  
Db 273 ATGCAAGTGTGGCCAAAGCGGGAAGATGGTGGGAGCCAGACACCGTTGGGATGAACTA 332  
Qy 374 TGAAGACTACATGGAAGAGAGACATATTCGCGCTCCAAATATGACAAACCAATGAACGAG 433  
Db 333 CGGAGCTACATGGAGGAGAGACATGATCCACCCCAACATGACACAGAGCGGAG 392  
Qy 434 AGTTATTTGCGCAGAGATCTTACGTTATGGAGACAGACCATGTACGGCAGTGGCTGA 493  
Db 393 AGTTATTCGTCGAGAGATCTTACGTTATGGAGTACAGACCATGTGCGGAGTGGCTGA 452  
Qy 494 GTGGGAGTGAAGGAGTATGGTCTTCCAGACGTGACATCTTGTGTTCCAGAACTATGA 553  
Db 453 GTGGGAGTGAAGGAGTATGGTCTTCCAGACGTCAACATCTTGTGTTCCAGAACTATGA 512  
Qy 554 TGGGAAGAGTGTGTAAATGACCAAGATGACCTTCCAGAGACTCAGCGCCAGCTATAA 613  
Db 513 TGGGAAGAGTGTGTAAATGACCAAGATGACCTTCCAGAGACTCAGCGCCAGCTATAA 572  
Qy 614 CGCAGATATCTCTCTGTCACACTACCTACCTCAGAGAGA-----GAGGAGCCAC 664  
Db 573 CGCGGACATCTCTCTCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 632  
Qy 665 TTTTATTTTCCAAATACATCAGTTTACCCAGAGCAACGCAAGAGATTAACAAAGAGCC 724  
Db 633 TTCAGATGATGTTGATAAAGCTTCAAAACCTCTCCACGCTTAAATGCAATGTAGAAAC 692  
Qy 725 AGATTTTACCTTATGAGCAGAGAGATCAGCGTGGAGAGTTCAGAGCATCCCA--C 781  
Db 693 AGATTTTACCTTATGAGCCCCCAGAGATCAGCTTGGACCGTTCAGCGCCACCCACGCC 752  
Qy 782 TCAGTCAAAAGTACCCCAACATCATCTTCAACAGTGTGCGCCAAAGACAGAGACAGCGTCC 841  
Db 753 CCAGTCGAAAGCTGCTCAACCATCTCTCTTCCACAGTGCCTCAAACTGAAAGACAGCGTCC 812  
Qy 842 TCAGTTAGATCCTTATCAGATCTTGGACCGACGAGCGCTTTCGAAATCCAGGAG 901  
Db 813 TCAGTTAGATCCTTATCAGATCTTGGACCAACAGTAGCGCGCTTCAAAATCCAGGAG 872  
Qy 902 TGGGAGATACAGCTATGCGAGTCTTCTTCTGAGCTTCTGTCGAGCAGCTCCCACTCCAA 961  
Db 873 TGGCCAGATCCAGCTTGGCGAGTCTCTCTGAGAGTCTCTGTCGAGCAGCTCCCACTCCAG 932  
Qy 962 CTGCATCACCTGGGAGGACAAATGGGAGTTCAGATGACAGACCTGATGAAGTGGC 1021  
Db 933 CTGCATCACCTGGGAGGACCAACAGCGGAGTTCAGATGACCGATCCCGACGAGTGGC 992

```

QY 1022 TCGGCGTTGGGAGAGAGAAAGCAAACTTAACATGAACATATGACAACTCAGCCGTC 1081
Db 993 CCGGCGTTGGGAGAGAGAGCAAACTTAACATGAACATATGACAACTCAGCCGTC 1052
QY 1082 ACTTCGCTACTACTATGACAAATATTAATGACTAAAGTTTATGCTAAACGCTATGCTTA 1141
Db 1053 CTTCCGTTACTACTATGACAAATATTAATGACTAAAGTTTATGCTAAACGCTATGCTTA 1112
QY 1142 CAAATTTGATTTCCACGGAATCGCTCAGGCGCTCCAGCCCTCACCTCCAGAAATCATCCAT 1201
Db 1113 CAAATTTGATTTCCACGGAATCGCTCAGGCGCTCCAGCCCTCACCTCCAGAAATCATCCAT 1172
QY 1202 GTACAAATACCCATCAGACCTCCCTTACATGATGTTCTTACATGACACACCCCGAGAGAT 1261
Db 1173 GTACAAATACCCATCAGACCTCCCTTACATGATGTTCTTACATGACACACCCCGAGAGAT 1232
QY 1262 GAACTTTGTAGTCTCCCATCCCTTGTCTTGGCCGTAACTCATCCAGCTTTTGTGTC 1321
Db 1233 GAACTTTGTAGTCTCCCATCCCTTGTCTTGGCCGTAACTCATCCAGCTTTTGTGTC 1292
QY 1322 CCTAATCATACTGGAATTCACCACTGGAGGATCTACCCGAATACAGGCTGCCAGC 1381
Db 1293 CCAAAACCCATCTGGAATTCACCACTGGAGGATCTACCCGAATACAGGCTGCCAGC 1352
QY 1382 TGCTCATATGCTTCCCATCTTGGCACCTACTACTAA 1418
Db 1353 CAGCCATATGCTTCTCATCTGGGCACCTACTACTAA 1389

RESULT 5
US-10-205-823-98
; Sequence 98, Application US/10205823
; Publication No. US20030108963M1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Gorbacheva, Bella
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Wonsley, Angela M.
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; THERAPY OF PROSTATE CANCER
; FILE REFERENCE: MRI-044
; CURRENT APPLICATION NUMBER: US/10/205,823
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 98
; LENGTH: 3166
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-205-823-98

```

Query Match 63.6%; Score 921; DB 15; Length 3166;  
 Best Local Similarity 81.1%; Pred. No. 2.5e-294;  
 Matches 1100; Conservative 0; Mismatches 245; Indels 12; Gaps 2;

```

QY 74 TATTAAGGAAGCATTATCACTGGTGAAGTGAAGACCAAGTCTCTTGTGTTGAGTGTGCTACGG 133
Db 289 TATCAAGGAAGCCTTATCACTGGTGAAGTGAAGACCAAGTCTCTTGTGTTGAGTGTGCTACGG 348
QY 134 ATGCCCCCCTTGGCAAGACAGAAATGACAGCTCTCTTCCAGTGAATATATGGGCAAAAC 193
Db 349 AACGCCACACCTGGCTAAGACAGAGATGACCGGCTCTCTTCCAGGACATATGGACAGAC 408
QY 194 ATCAAGATGAGCCCGCGCTTCCCGACGAGGACTGTTTATCACAGCCCCCGCCAGAGT 253
Db 409 TTCCAAGATGAGCCCGCGCTCTCTTCCAGGAGTTGGCTGTCTCAACCCCGCCAGGAGT 468
QY 254 TACCATTAAGATGAGGCTGTAACCCAAACAGGTTAATGGGTCAAGAAATTCACCTGATGA 313
Db 469 CACCATCAAAATGAATGAACCTTAGCCAGGTTGAATGGCTCAAGAAATCTCTCTGATGA 528
QY 314 CTCAGCGTGGCAAAAGGAGGAAATGGTTAGCAGTTTACAGCAATATGTTGGGATGAACATA 373
Db 529 ATCAGTGTGGCAAAAGGCGGGAAGATGTTGGGACCCCGACACACCTTTGGGATGAACATA 588
QY 374 TGAAGCTACATGGAAGAGAGAGCAATATTCGGCTTCAAAATATGACAAACCAATGAAG 433
Db 589 CGCAGCTACATGGAAGAGAGAGCAATATTCGGCTTCAAAATATGACAAACCAATGAAG 648
QY 434 AGTTATTGTGCCAGCAGATCTTACGTTATGGAGACAGACCACTGTACGGCAGTGGGTGGA 493
Db 649 AGTTATTGTGCCAGCAGATCTTACGTTATGGAGTACAGACCATGTGCGCAGTGGGTGGA 708
QY 494 GTGGCAGTGAAGAGTATGCTTCTTCCAGCTGGGACATCTTGTGTTTCCAGAACTATGA 553
Db 709 GTGGGCGGTGAAGAAATATGGCTTCCAGACGTTCAACATCTTGTATTCAGAACTATGA 768
QY 554 TGGGAAGAGTGTGTGAATAATGACCAAGATGACTTCCAGAGACTCACGCCGAGCTATAA 613
Db 769 TGGGAAGAGTGTGTGAATAATGACCAAGATGACTTCCAGAGACTCACGCCGAGCTATAA 828
QY 614 CGCAGATATCTCTCTGTCACCTACACTACTCTCAGAGAGA-----GAGAGCCAC 664
Db 829 CGCCGACATCTCTCTCACATCTCCACTACTCTCAGAGAGACTCTCTCTTCCACATTTGAC 888
QY 665 TTTTATTTTCCAAATACATCAGTTTACCCAGAGCAACGCAAGAAATAACCAACAGGCC 724
Db 889 TTCAGATGATGTTGAATAAGCCCTTACAAACTCTCCAGGTTAATGCATGCTTGAACAC 948
QY 725 AGATTTACCTTTATGAGCAAGCAGAGAGATCAGCGTGGAGAGTCAAGCCATCCCA---C 781
Db 949 AGATTTACCATATGAGCCCCCAGGAGATCAGCGTGGAGAGTCAAGCCATCCCAAGGCC 1008
QY 782 TCAGTCAAAAGCTACCCACCATCATCTTCAACAGTGGCCCAAAACAGAGACCCAGGTC 841
Db 1009 CCAGTCGAAAGCTGCTCAACCATCTCTTCCAGTGGCCCAAAACAGAGACCCAGGTC 1068
QY 842 TCAGTTAGATCTCTTATCAGATCTTGGACCGACAGCAGCCGCTTTGCAAAATCCAGGAG 901
Db 1069 TCAGTTAGATCTCTTATCAGATCTTGGACCAACAAGTAGCCGCTTTGCAAAATCCAGGAG 1128
QY 902 TGGGCGAGATACAGTATGAGGTTCTTCTGAGGTTCTTGGAGAGTCTTCCAGCTCCAACTCAA 961
Db 1129 TGGCCAGATCCAGCTTGGCAGTCTCTCTGAGGCTCTCTTGGAGAGTCTTCCAACTCCAG 1188
QY 962 CTGCATCACCTGGGAGGCGACAAATGGGAGTTCAAGATGACAGACCCCTGATGAAGTGGC 1021
Db 1189 CTGCATCACCTGGGAGGCGACCAACGGGAGTTCAAGATGACAGGATCCCGCAGGTTGGC 1248
QY 1022 TCGGCGTTGGGAGAGAGGAAAGCAACCTTAACATGAATATGACAACTCAGCCGTC 1081
Db 1249 CCGGCGCTGGGAGAGAGGAAAGCAACCTTAACATGAATATGACAACTCAGCCGTC 1308
QY 1082 ACTTCGCTACTACTATGACAAATATTAATGACTAAAGTTTATGTTGTTAAACGCTATGCCTA 1141
Db 1309 CTTCCGTTACTACTATGACAAAGAAATCATGACCAAGTCCATGGAGGCTTACGCTTA 1368

```



```
US-10-087-192-1124
; Sequence 1124, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 52945200122
; CURRENT APPLICATION NUMBER: US/10/087,192
; PRIOR FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1124
; LENGTH: 1297
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-087-192-1124

Query Match      52.7%; Score 763.2; DB 13; Length 1297;
Best Local Similarity 75.3%; Pred. No. 4,1e-242;
Matches 1008; Conservative 0; Mismatches 288; Indels 42; Gaps 3;

QY 80 GGAAGCATTATCAGTGTGAGTGAAGACCAAGTCTCTTTGAGTGTGCTTACGGATCGCC 139
DB 1 GGAGGCCCTTGTGAGTGTGAGGAGGACCAAGTCACTATTGAGTGTGCTTACGGAGCGCC 60
QY 140 CCACCTTGCAGAGACAGAAATGACAGCTCTCTTCCAGTGAATATGGGCAAAATCAGA 199
DB 61 ACACCTGGCTAAGACAGAGATGACCGCATCTCTTCCAGTGAATATGGGCAAGATCCAA 120
QY 200 GATGAGCCCGCGCTTCCACAGAGAGTGGTATACAGCCCCCGCCAGAGATTACCAT 259
DB 121 GATGAGTCCCAAGTCTCCTCAGCAGGACTGGCTGTCTCAAGCCCGCCAGGCTACCAT 180
QY 260 TAAGATGAGTGTAAACCAACAGGTTAATGGTCAAGGANTTCACTGATCACTGTCAG 319
DB 181 CAAGATGAGTGCACACCTAGTCAAGTGAATGGTTCAGGAACCTCACTGATGAGTGCAG 240
QY 320 CGTGCGAAAAGAGGAGAAATGGTTAGCAGTTTCAGACAAATGTTGGGATGAACATATGAAG 379
DB 241 TGTGAACAAAGGTGGAGAGATGGTGGCAGCCCGGATGCTGTGGGATGAGCTACGGCAG 300
QY 380 CTACATGGAAGAGAGATATATCCGCTCCAAATATGACAAACCAATGAACGAGAGTTAT 439
DB 301 CTACATGGAAGAGAGATGATGGCGCTCCCAATATGACCAACAAATGAGCGCAGAGTGAT 360
QY 440 TGTGCCAGCAGATCCTACGTTATGGAGCAGACACATGTACGGCAGTGGCTGGAGTGGGC 499
DB 361 CGTCCCTGCAGATCCTACTCTGTGAGCAGACAGACATGTCCGACAGTGGCTGGAGTGGGC 420
QY 500 AGTGAAGAGATGGTCTTCAGACGTTGGACATCTTGTGTCAGAACATTTGATGGGAA 559
DB 421 GGTGAAAGAAATATGGCTCTCTCGATGTGGAGCTTACTATTTCAGAAATATCGATGGGAA 480
QY 560 AGAGTGTGTAATGACCAAGATGACTTCCAGAGACTCAGCGCAGCTATAACGCAGA 619
DB 481 GGAGCTGTGCAAGATGACAAAGGATGACTTCCAGCGGCTCAGCGGAGCTACATGCCGA 540
QY 620 TATCCTCTGTGACACCTTACCTACCTCAGAGAGAGAGGAGCCACTTTTATTTTCCAAA 679
DB 541 CATTTCTCTCTCACATCTCCACTACCTCAGAGAGA-----CTCCCGCTTC 584
QY 680 TACATCAGTTTACCAGAGACGCAAGATTAACCAAGGCCAGATTTACTTATGA 739
DB 585 CACATCTGACTTCGATGACCTTGATAGGCTTTTACAAA-----CTCTCCAGGTTAAT 639
QY 740 GCAAGCGAGGAGATCAGCGTGGAGAGTTCACAGCCATCCCACTCAGTCAAAAGCTACCCA 799
```

## RESULT 8

```
US-10-052-482-195
; Sequence 195, Application US/10052482
; Publication No. US2004007226A1
; GENERAL INFORMATION:
; APPLICANT: Engelhard, Eric
; APPLICANT: Morris, David
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71087/RMS/DCF
; CURRENT APPLICATION NUMBER: US/10/052,482
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 241
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 195
; LENGTH: 1359
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-052-482-195
```

```
Query Match      39.4%; Score 570.2; DB 12; Length 1359;
Best Local Similarity 66.8%; Pred. No. 6.3e-178;
Matches 917; Conservative 0; Mismatches 423; Indels 33; Gaps 6;
```



QY 53 ATGGCAAGCACTATTAAAGGAGCATTATCAGTGTGAGTGAAGACCAAGTCTCTTTTGGAG 122  
 Db |||||  
 QY 1 ATGGAGGGAGCTATTAAAGAGGCTCTGTCTGTGTGAGTGAAGTCAAGTCCCTTTTGGAT 60  
 Db |||||  
 QY 123 TGTGCTACGGATCG---CCCAACCTTGCAGAGAGCAGAAATGACAGAGCTCTCTTCCAGT 179  
 Db |||||  
 QY 61 TCAGCATACGGAGCGGAGCCCATCTCCCAAGGCAGATATGACTGTCTCGGAGGCTCT 120  
 Db |||||  
 QY 180 GAAATATGGGCAACATCAAGATGAGCCCGCGGTTCCTCCAGAGAGAGTGGTATACAG 239  
 Db |||||  
 QY 121 GACTACGGGAGCGCCCAAAAATCAACCCCTGCGCACCGCAGCAGGAGTGGATCAACAG 180  
 Db |||||  
 QY 240 CCCCCGGCCAGAGTTACCAATTAAGATGAGTGTAAACCCCAACACAGGTTAATGGGTCAAGG 299  
 Db |||||  
 QY 181 CC-----AGTCAGAGTCAATGTCAAGCGGAGTATGACCATGAATGGATCCAGG 231  
 Db |||||  
 QY 300 AATTCACCTGATGCTGAGCGGTGCGCAAAAGAGGAGGAAAATGGTATGAGTTACAGAAAT 359  
 Db |||||  
 QY 232 GAGTCTCCGGTGGAGTGCAGTGTGAGCAATGTAAACAGTGTGGCGGAGGCGAAGCC 291  
 Db |||||  
 QY 360 GTTGGGATGAACTATGGAAGCTACATGGAAGAGAGCA---TATTCGGCTCCCAATATG 416  
 Db |||||  
 QY 292 AACCCCATGAACATAAATAGCTACATGATGAGAGAAACGGCCCCCTCTCTCCCAATG 351  
 Db |||||  
 QY 417 ACAACCAATGACGAGAGTATTATGTCAGCAGATCTCTACGTTATGAGCAGACAGCAAT 476  
 Db |||||  
 QY 352 ACCACCAAGCAAGAGAGTATTGTGCTGACAGACCCCACTGTGCACAGAGGAGC 411  
 Db |||||  
 QY 477 GTACGGCAGTGGCTGAGTGGCAGTGAAGAGATGTCTTCAGAGCTGGACATCTTG 536  
 Db |||||  
 QY 412 GTTCAGAGTGGCTGGAGTGGCTATAAAGAAATACGANTGATGGAGATTGACATCTCC 471  
 Db |||||  
 QY 537 TTGTTCCAGAACTATGATGGGAAAGAGTTGTGTAAATGACCAAGATGATCTCCAGAG 596  
 Db |||||  
 QY 472 TTCTCCAGAACATGATGGCAGAGAAATGTGTAAATGAACAAGGAGGACTTCTCCGA 531  
 Db |||||  
 QY 597 CTCAGCGCAGATTAAGCGAGATATCTCTGTGACACCTACACTACCTCAGAGAGAG 656  
 Db |||||  
 QY 532 GCCACTCCGCTACACACAGAGTGTGTGTGCGACCTCAGTTACTCTCAGGGAAG- 590  
 Db |||||  
 QY 657 GGAGCCACTTTTATTTTCCAAATACATCAGTTTACCCAGAGCAAGCAAGAAATAACA 716  
 Db |||||  
 QY 591 -----TTCACTGCTGCTTATAACAACTCCCATACAGACCAAGTCTCAGACTGAT 645  
 Db |||||  
 QY 717 ACAAGCCAGATTTTACCTTATGAGCAAGCAGGAGAGATCAGGTGAGAGTCAACA---GC 773  
 Db |||||  
 QY 646 GTCAAGGAAGACCCCTCTTATGACTCTCTCAGGAGAGGAGCATGGAACAATAATGAAC 705  
 Db |||||  
 QY 774 CATCCCACTCAGTCAAGAGCTACCAACCATCTTCAACAGTCCCAACCAAGAGAC 833  
 Db |||||  
 QY 706 TCTGGCTCAACAAAGTCTCTCTCTGGAGATACAGACCATGGGCAAGACACTGAG 765  
 Db |||||  
 QY 834 CAGCGTCTCAGTTAGATCCTTATCAGATTCTTGGACCGACAGCAGCGCTTTGCAAAAT 893  
 Db |||||  
 QY 766 CAGCGCCCCAGCAGATCCTTATCAGATCTTGGGCGCAACAGCAGCGGCTAGCAAAAC 825  
 Db |||||  
 QY 894 CAGGAGTGGCAGATACAGATATGAGATCTCTTCTGAGAGTCTCTGCGAGAGCTCC 953  
 Db |||||  
 QY 826 CTTGGAGTGGCAGATCAGCTGTGGCAGTTCTCTTGGAACTACTGTCCGAGAGCGCC 885  
 Db |||||  
 QY 954 AACTCCAACTGATCACTCTGGAGGCGCAAAATGGGGAGTTCAAGATGACAGACCTCTGAT 1013  
 Db |||||  
 QY 886 AACCCAGCTGTATCACTCTGGAGGGGACCAAGCGGAGTTCAAAATGACGAGCCCTGAT 945  
 Db |||||  
 QY 1014 GAAATGGCTGGCGTGGGAGAGAGGAAAGCAACCTAACATGAATATGACAAACTC 1073  
 Db |||||  
 QY 946 GAGTGGCCAGCGCTGGGAGAGCGGAGAGCAAGCCCAACATGAATATATGCAAGCTG 1005  
 Db |||||  
 QY 1074 AGCCGTCGACTTCGCTACTACTATGACAAAATATTATGACTAAAGTTTCATGGTAAACGC 1133  
 Db |||||  
 QY 1006 AGCCGGGCCCTCCGATACTACTATGACAAACATTTAGCAAAAGTGCATGGCAAGG 1065  
 Db |||||

QY 1134 TATGCTCAAAATTTGATTTCCAGGAATCGTCCAGCCCTCAGCCTCACCTCCAGAA 1193  
 Db |||||  
 QY 1066 TATGCTCAAAATTTGATTTCCAGGAATCGTCCAGCCCTCAGCCTCACCTCCAGAG 1125  
 Db |||||  
 QY 1194 TCATCCATGTACAAATATACCATCAGACCTCCCTACATGAGTTCTTACCATCAGACCC 1253  
 Db |||||  
 QY 1126 ACATCCATGTACAAATATCCCTCTGATATCTCTACATGCTTCTTACCATGCCATCA 1185  
 Db |||||  
 QY 1254 CAGAGATGAATTTGTAGTCCCATCCCCCTGCTTTTCCCGGTAACTCATCCAGCTTT 1313  
 Db |||||  
 QY 1186 CAGAGGTGAATTTGTCCCTGCTCACCCTCTCCATGCTGTCCCTCTCCAGCTTC 1245  
 Db |||||  
 QY 1314 TTTGCTGCCCTTAATCCATATCTGGAATTCACCAACTGGAGGATCTTACCCCAATACAGG 1373  
 Db |||||  
 QY 1246 TTTGGAGCAGCATCAATATCTGAGCTCCCTCCCACTGTGGATCTATCCAAACCCAGT 1305  
 Db |||||  
 QY 1374 CTGCCAGCTGCTCAT- - - - -TGCTTCCCATCTTGGCACTTACTACTA 1417  
 Db |||||  
 QY 1306 GTCCCCCGCCATCTTAACACCCACGTCGCTTACACTTAGGAGCTACTACTA 1358  
 Db |||||

RESULT 9  
 US-10-052-482-194  
 ; Sequence 194, Application US/10052482  
 ; Publication No. US20040072264A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Engelhard, Eric  
 ; APPLICANT: Morris, David  
 ; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER  
 ; FILE REFERENCE: A-71087/RMS/DCF  
 ; CURRENT APPLICATION NUMBER: US/10/052,482  
 ; CURRENT FILING DATE: 2002-08-15  
 ; PRIOR APPLICATION NUMBER: US 09/747,377  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: US 09/798,586  
 ; PRIOR FILING DATE: 2001-03-02  
 ; NUMBER OF SEQ ID NOS: 241  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 194  
 ; LENGTH: 1729  
 ; TYPE: DNA  
 ; ORGANISM: Mus musculus  
 US-10-052-482-194

Query Match 39.4%; Score 570.2; DB 12; Length 1729;  
 Best Local Similarity 66.8%; Pred. No. 7.4e-176; Indels 33; Gaps 6;  
 Matches 917; Conservative 0; Mismatches 423;

QY 63 ATGGCAAGCACTATTAAAGGAGCATTATCAGTGTGAGTGAAGACCAAGTCTCTTTTGGAG 122  
 Db |||||  
 QY 245 ATGGACGGGACTATTAAAGGAGCTCTGTCTGTGTGAGTGAAGTCAAGTCCCTTTTGGAT 304  
 Db |||||  
 QY 123 TGTGCTACGGATCG---CCCAACCTTGCAGAGAGAGAAATGACAGCTCTCTTCCAGT 179  
 Db |||||  
 QY 305 TCAGCATACGGAGCGGAGCCCATCTCCCAAGGAGATATGACTGCTTCGGGGAGTCT 364  
 Db |||||  
 QY 180 GAATATGGCAACATCAAGATGAGCCCGCTTCCCAAGAGTCTTATCACAG 239  
 Db |||||  
 QY 365 GACTACGGGAGCGCCCAAAATCAACCTTCCCAAGAGAGTGGATCAACAG 424  
 Db |||||  
 QY 240 CCCCCGGCAGAGTTACCAATTAAGATGAGTGTAAACCAACAGGTTAATGGTCAAG 299  
 Db |||||  
 QY 425 CC-----AGTCAGAGTCAATGTCAAGCGGAGTATGACCAATGAATGATCCAGG 475  
 Db |||||  
 QY 300 AATTCACCTGATGACTGACGCTGGCAAGAGGAGAAATGGTATGAGTTCAGACAT 359  
 Db |||||  
 QY 476 GAGTCTCCGGTGGAGTGTGAGTGTGAGCAATGTAAACAGCTGTGGCGGAGCGAAGCC 535  
 Db |||||  
 QY 360 GTTGGGATGAACCTATGGAAGTACATGGAAGAGAGCA---TATTCGGCTTCAATATG 416  
 Db |||||  
 QY 536 AACCCCATGAACCTATAATAGTACATGATGAGAGAGCGGCCCTCTCTCCCAACATG 595  
 Db |||||  
 QY 417 ACAACCAATGAAGAGAGTATTATGTCAGCAGATCTTATGTCAGCAGATCTTATGAGCAGACAGACCAT 476  
 Db |||||



596 ACCACCAACGACGAGGATCATTTGGCTGCGACACCCACACCTGTGGACACAGGAGCAC 655  
477 GTACGGAGTGGCTGGAGTGGAGAGTATGGCTTCCAGACGTTGGACATCTTG 536  
656 GTTCGACAGTGGCTGGAGTGGAGTATGAAGAAATACGATTTGAGATGACACTTCC 715  
537 TTGTTCCAGAACTGTATGGGAAAGAGTTGTGTAATAATGACCAAGATGACTTCCAGAGA 596  
716 TTCTTCCAGAACATGGATGGCAAGGAATTTGTGAATATGAACAAGGAGGACTTCTCCGA 775  
597 CTCAGCGGAGCTATAAGCAGATATCTCTCTGTGTCACACCTACACTACCTCAGAGAGAGA 656  
776 GCCACCTCCGCTCAACACAGAAAGTGTGTGTGCGACCTCAGTTACCTCAGGGAAG- 834  
657 GGAGCCACTTTATTTTCCAAATACATCATGTTTACCCAGAGCAACGCAAGAAATAACA 716  
835 -----TTCACTGTGGCTATACACAACTCCCATACAGACAGTCTCTCAGACTGAAT 889  
717 ACAAGGCCAGATTTACCTTATGAGCAAGCGAGGAGATCAGCGTGGACGAGTCAACA---GC 773  
890 GTCAAGGAAGACCTTCTTATGACTCTGTCTAGGAGAGGAGCATGGAACAATAATATGAAC 949  
774 CATCCCACTAGTCAAAAGCTACCAACCATCTCTCTTGGAGGATCAAGACCATGGGCAAGAACACTGAG 1009  
950 TCTGGCTCAACAAAGTCTCTCTCTTGGAGGATCAAGACCATGGGCAAGAACACTGAG 1009  
834 CAGCGTCTCTAGTTAGATCTCTTATCAGATTTCTGAGCGGACCGACGAGCGCTCTTGCAGAAAT 893  
1010 CAGCGGCCAGCAGATCTCTTATCAGATCTCTGCGGCCACACGACGCGCTAGCAAAAC 1069  
894 CGAGGAGTGGGAGATACAGCTATGGCAGTTCTCTACTGGAGCTTCTGTGGAGAGTCTCC 953  
1070 CTTGGAGTGGGAGATCCAGCTGTGGCAGTTCTCTCTGGAACACTGTCTCGACAGCGCC 1129  
954 AACTCCAACTCATCACCTGGGAGGACAAATGGGAGTTCAAGATGACAGACCTCTGAT 1013  
1130 AACGCCAGCTGTATCACCTGGGAGGAGACCAAGCGGAGTTCAAAATGACGGACCTGAT 1189  
1014 GAAGTGGCTCGGCTTGGGAGAGAGAAAGCAAACTTAATGAACTATGACAACTC 1073  
1190 GAGTGGCCAGCGCTGGGAGAGCGGAAGAGCAAGCCCAACATGAATATGACAACTG 1249  
1074 AGCGTGACACTTCGCTACTACTATGACAAATAATATGACTAAAGTTCAATGGTAAACGC 1133  
1250 AGCGGGCCCTCCGATACTATGACAAATAATATGACTAAAGTTCAATGGTAAACGC 1309  
1134 TATGCTTACAAATTTGATTTCCACGGAATCGCTCAGGCGCTTCCAGCTTCACTCCAGAA 1193  
1310 TATGCTTACAAATTTGATTTCCATGGCATTTGCCAGGCGCTTCCAGGCGATCCCAACAGAG 1369  
1194 TCATCCATGTACAAATACCAATCAGACCTCCCTTACATGATTTCTTACCATGACACCCC 1253  
1370 ACATCCATGTACAAATACCAATCAGACCTCCCTTACATGATTTCTTACCATGACACCCC 1429  
1254 CAGAGATGAACTTTGTAGCTCCCATCCCTCTGTTTGGCCGCTAACTCATCCAGCTTT 1313  
1430 CAGAGAGTGAATTTGTCCGCTCTACCCATCCCTCCATGCTGCTCTCCAGCTTC 1489  
1314 TTGTGTCCTTAAATCCATATGAGAAATTCACCACTGAGGAGCTTACCCCAATACAGG 1373  
1490 TTTGGAGCAGCATCAATATCTGAGCTCCCTTCCCTTCTGCTGGGATCTATCCAAACCCAGT 1549  
1374 CTGCCAGTGTCTATA-----TGCTTCCCATCTTGGCAGCTACTACTA 1417  
1550 GTCCCCCGGCTCTTACACCCACGCTGCTTACACTTTAGCGAGCTACTACTA 1602

RESULT 10  
US-10-052-482-198  
; Sequence 198, Application US/10052482  
; Publication No. US200400722641  
; GENERAL INFORMATION:

APPLICANT: Engelhard, Eric  
; APPLICANT: Morris, David  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER  
; FILE REFERENCE: A-71087/RMS/DCF  
; CURRENT APPLICATION NUMBER: US/10/052,482  
; CURRENT FILING DATE: 2002-08-15  
; PRIOR APPLICATION NUMBER: US 09/747,377  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 09/798,586  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 241  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 198  
; LENGTH: 1359  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-052-482-198  
  
Query Match 38.9%; Score 562.2; DB 12; Length 1359;  
Best Local Similarity 66.4%; Pred. No. 2.9e-175;  
Matches 912; Conservative 0; Mismatches 428; Indels 33; Gaps 6;  
  
QY 63 ATGGCAAGCACTATTAAAGGAAGCATTTATCAGTGGTGAAGAGACCAAGTCTTGTGTTGAG 122  
DB 1 ATGGACGGGACCTTTAAGGAGGCTCTGTGCTGTGGTGGAGGAGCAACATGCTCTTTGAC 60  
QY 123 TGTGCTACGAGTCG---CCCCACCTTGCAGAGAGAGAGAAATGACAGCTCTCTCTTCCAGT 179  
DB 61 TCAGCGTACGAGCGCGAGCCCATCTCCCAAGGCGACATGACTGCTCTCGGGAGTCTCT 120  
QY 180 GAATATGGGCAACATCAAGATGAGCGCGCTTCCCGACGAGCTGCTTATCACAG 239  
DB 121 GACTACGGGAGCGCCCAAGATCAACCCCTCCACACAGAGGAGTGGATCAATCAG 180  
QY 240 CCCCCGGCAGAGTTACCATTAAGATGGAGTGTAAACCAACAGGTTAATGGTCAAG 299  
DB 181 CC-----AGTAGGGGTCAACGTCAAGCGGGAGTATGACCACATGAATGGATCCAG 231  
QY 300 AATTCACTCTGATGACTGCGAGCGTGGCAAGAGGAGGAAATGTTAGCAGTTTCAGACAAT 359  
DB 232 GAGTCTCCCGTGGAGTGGAGCGCTTAGCAATGACGACGCTGCTGGCGGAGCGAGTCC 291  
QY 360 GTTGGATCAACTATGGAAGCTACATGGAAGAGAGCAT---ATTCGCGCTCCAAATATG 416  
DB 292 AACCCCATGACTACAAACAGCTATATGGACGAGAAAGATGGCGCCCTCTCTCCCAATG 351  
QY 417 ACACCAATGACGAGAGTATTTGTCGAGCAGATCTCTACGTTATGGAGCAGACCAT 476  
DB 352 ACCACCAAGAGAGAGTCTGTCCTCCCGCAGACCCCACTGTGGACAGAGCAT 411  
QY 477 GTACGGCAGTGGCTGGAGTGGGAGTGAAGAGTATGGTCTTCCAGACGTTGACATCTTG 536  
DB 412 GTGAGCAATGGCTGGAGTGGGCAATAAAGGAGTACAGCTTGATGGAGATCGACACATCC 471  
QY 537 TTGTTCCAGAACTATGATGGGAAAGTGTGTGTAATAATGACCAAGATGACTTCCAGAGA 596  
DB 472 TTTTTCAGAACTATGATGGGAAAGTGTGTGTAATAATGACCAAGAGGAGTCTCTCCGC 531  
QY 597 CTCACGCCGAGCTATAACCGCAGATATCTCTCTGTCACACTACACTACCTCAGAGAGAGA 656  
DB 532 GCACCAACCTCTACAAAGGAGTGTGTGTGTCACACTCAGTTACCTCAGGGAAG- 590  
QY 657 GGAGCCACTTTTATTTTCCAAATACATCAGTTTACCCAGAGCAACGCAAGAAATAACA 716  
DB 591 -----TTCACTGTGGCTTATAATAAAGCTCTCCACACCGCAACATCTCTCAGATTGAGT 645  
QY 717 ACAAGGCCAGTATTAACCTTATGAGCAAGCGAGGAGATCAGCGTGGAGGAGTCAAGCCAT 776  
DB 646 GTCAAGGAAGACCTTCTTATGACTCAGTCAAGAGGAGCTTTGGGCGCATATACATGAAT 705  
QY 777 CC---CACTCAGTCAAAAAGCTACCCAAACCATCATCTTCAACAGTGGCCCAAAACAGAGAC 833  
DB 706 TCTGGCTCAACAAAGTCTCTCCCTTTGGAGGGGACAAACGATCAGTAAGATAACAGAG 765

```

QY 834 CAGCGTCTCAGTTAGATCCCTTATCAGATCTTGGACCGACAGAGCCGCTTGTGCAAT 893
DB 766 CAACGGCCCCAGCCAGATCCGATATCAGATCTGGGCCCCGACAGCATCGCTTAGCCAAC 825
QY 894 CCAGGAGTGGGAGATACAGTATGCGATTCCTACTCTGAGCTTCTGCGAGCTCC 953
DB 826 CTTGGAAGCGGCAGATCCAGCTGTGGCAATTCCTCTGAGCTGCTTCCGACAGGCC 885
QY 954 AACTCCAACTGCATCCTTGGAGGGGACAAATGGGGAGTTCAAGATGACAGACCTGAT 1013
DB 886 AACGCCAGCTGTATCACTGGAGGGGACCAACGGGGAGTTCAAAATGACGGACCCCGAT 945
QY 1014 GAAGTGGCTGGGCTTGGGAGAGAGGAAAGCAAACTTAACATGAATGACAAACTC 1073
DB 946 GAGTGGCCAGCCCTGGGGGAGCGGAAAGAGCCCAACATGAATACGACAGCTG 1005
QY 1074 AGCGTGCATCTCCTACTACTATGACAAATAATATGATAAAGTTTCATGTAACGC 1133
DB 1006 AGCGGGCCCTCCGTTTACTATGATAAAACATTAATGACCAAGTGCACGGCAAAAGA 1065
QY 1134 TATGCTACAAATTTGATTTCCACGGAATCGCTCAGGCCCTCCAGCCTCACCCTCCAGRA 1193
DB 1066 TATGCTACAAATTTGATTTCCACGGAATTCGCGGATTCGCCAGGCTCTGACGCCACATCCGACGGAG 1125
QY 1194 TCATCCCATGTACAAATACCCATCAGACCTCCCTTACATGAGTTCTTACCATGACACACCCC 1253
DB 1126 TCGTCCATGTACAGTACCCCTTCTGACATCTCTTACATGCTTCTTACCATGCCACCCAG 1185
QY 1254 CAGAGATGAACTTTGATCTCCCATCCCTCCCTGCTTGGCGGTAACTCATCAGCTTT 1313
DB 1186 CAGAGGTGAATTTGCTCCCTCCCATCCATCCCTCCATGCTGCTTCTCCAGCTTC 1245
QY 1314 TTGTGTCCTTAAATCCATACCTGGAATTCACCACTGGAGGCACTTACCCCAATACCA-- 1371
DB 1246 TTGGAGCGCATCACAACTATGACCTCCCGACGGGGGATCTACCCCAACCCCAAC 1305
QY 1372 -----GCTGCGAGCTGCTCATATGCTTCCATCTTGGCACCCTACTACTA 1417
DB 1306 GTCCCCGCCATCTTAACACACCGCTGCTTCAACATTAGCGAGCTACTACTA 1358

```

RESULT 11

```

US-10-052-482-197
; Sequence 197, Application US/10052482
; Publication No. US20040072264A1
; GENERAL INFORMATION:
; APPLICANT: Engelhard, Eric
; APPLICANT: Morris, David
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71087/RMS/DCF
; CURRENT APPLICATION NUMBER: US/10/052,482
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 241
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 197
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-052-482-197

```

```

Query Match 38.9%; Score 562.2; DB 12; Length 2957;
Best Local Similarity 66.4%; Pred. No. 4.9e-175;
Matches 912; Conservative 0; Mismatches 428; Indels 33; Gaps 6;
QY 63 ATGGCAAGCATTATTAAGGAGCATATCAGTGGTGAAGTGAAGCAGCTCTCTTTGAG 122
DB 173 ATGGACGGGACTATTAGAGAGCTCTGTCGGTGGTGGTGGAGCAGCAGCTCTCTTTGAC 232

```

```

QY 123 TGTGCTACGGATCG---CCCCACCTTGCAGAGAGAGAAATGACAGACCTCTCTTCCAGT 179
DB 233 TCAGGTACGGAGCGGAGCCCATCTCCCAAGGCGGACATGATGCTCTCGGGAGTCTT 292
QY 180 GAATATGGGCAAAACATCAAGATGAGCCCGCGCTTCCCCAGCAGGAGTGTGTATCAG 239
DB 293 GACTACGGGCGAGCCGCCACAGATCAACCCCTCCACCCACAGCAGGAGTGTATCAATCAG 352
QY 240 CCCCCGGCCAGAGTTACCATTAAGATGGAGTGAACCCAAACCCAGGTTAAATGGGTCAAGG 299
DB 353 CC-----AGTGAGGTCAACGCTCAAGCGGGAGTATGACCATCATGATGATCCAGG 403
QY 300 AATTCCATCTGATGACTGAGCGGTGGCAAAAGAGGAGGAAATGGTTAGCAGTTACAGCAAT 359
DB 404 GAGTCTCCGTTGAGCTGAGCGGTAGCAATTCAGCAAGCTGGTGGCGAGGAGGAGTCC 463
QY 360 GTTGGATGAACTATGGAAGCTACATGGAAGAGAGCAT---ATTCCGCTCCCAATATG 416
DB 464 AACCCCATGAACTACACAGCTATATGACGAGAGAAATGGCCCCCTCTCTCCCAATG 523
QY 417 ACAACCAATGAACGAAGAGTTATTGTGCGAGCAGATCTTACGTTATGGAGCAGACCAT 476
DB 524 ACACACAGAGAGAGAGTATCATGCTCCCGCAGACCCACACATGTGGACAGAGCAT 583
QY 477 GTACGGAGTGGCTGGAGTGGGAGTGAAGAGTATGCTTCCAGAGCTGACATCTTG 536
DB 584 GTGAGCAATGCTGGAGTGGGCCATAAAGGAGTACAGCTTGAAGAGATGCACATCC 643
QY 537 TTGTTCCAGAACATGATGGGAAGAGTTGTGTAATGACCAAGAGTACTTCCAGAGA 596
DB 644 TTTTTCAGAACATGATGGGAGAGTGTGTTGTTGTCACACCTCAGTTACTCAGGAGAAAG- 703
QY 597 CTCAGCGGAGTATAACGAGATATCCTCTGTGTCACACTACACTACTCAGAGAGAGA 656
DB 704 GGCACACCTCTACACACGGAAGTGTGTTGTCACACCTCAGTTACTCAGGAGAAAG- 762
QY 657 GGAGCCACTTTTATTTTCCAAATACATCAGTTTACCAGAGCAACGCAAGAGATAACA 716
DB 763 -----TTCACTGCTGGCTATATATACCTCCACACCGACCAATCTCTCAGATTGAT 817
QY 717 ACAAGCCAGATTTACCTTATGAGCAAGGAGAGATCAGCGTGAAGAGTACAGCAT 776
DB 818 GTCAAGAGAGACCTTCTTATGACTCAGTCAAGAGAGAGCTTGGGGCAATTAACATGAT 877
QY 777 CG---CACTCAGTCAAAAGCTACCCACCATCATCTTCAACAGTGCCTCCCAACAGAGAC 833
DB 878 TGTGGCTCAACAAAGTCTCTCCCTTGGAGGGGACAAACGATCAGTAAAGATACAGAG 937
QY 834 CAGCGTCTCAGTTAGATCTTTATCAGATTTTGGACCGGACAGAGCGCTTGTGCAAT 893
DB 938 CAACGCCCCAGCAGATCCGATCAGATCTCTGGGCGGACCCAGCAGTGCCTTAGCCAC 997
QY 894 CCAGGAGTGGGAGATACAGCTATGCGAGTTCTTACTGGAGCTTCTGTCGACAGCTCC 953
DB 998 CTTGGAAGCGGCGAGATCCAGCTGTGGCAATTTCTCTGGAGCTGCTCTCCAGCAGCGCC 1057
QY 954 AACTCCAACTGCATCACCTGGGAGGGGCAAAATGGGGAGTTCAAGATGACAGACCTGAT 1013
DB 1058 AACGCGAGCTGTATCACTGGGAGGGGACCAACGGGGAGTTCAAATGACGACACCCGAT 1117
QY 1014 GAAGTGGCTGGGTTGGGAGAGAGGAAACCAACCTAAACATGAACTATGACAACTC 1073
DB 1118 GAGTGGCGAGCGCTGGGGCGAGCGGAAAGCAAGAGCCCAACATGAATACGACAGCTG 1177
QY 1074 AGCGGTGCACTTTCGCTTACTTACTATGACAAATAATTTATGACTTAAAGTTTATGTTAAACGC 1133
DB 1178 AGCGGGCCCTCCGTTTACTTATGATAAAACATTTATGACCAAAAGTGCACGGCAAAAGA 1237
QY 1134 TATGCTCAAAATTTGATTTCCAGGAATGCTCAGGCCCTCAGCCTCAGCTCCAGAA 1193
DB 1238 TATGCTTACAAATTTGACTTCCACGCAATTTCCAGGCTCTCGAGCCATCCGACCGAG 1297
QY 1194 TCATCCATGTACAAATFACCCATCAGACCTCCCTCTACATGATGATTCTTACCATGACACCCC 1253

```

Db 1298 TCGTCAAGTAAAGTACCCCTTCTGACATCTCTACATGCTTCTTACCAGTCCACAG 1357  
Qy 1254 CAGAAGATGAATTTGTAGCTCCCATCCCTCTCTTGGCCGTAACCTCATCCAGCTTT 1313  
Db 1358 CAGAAGGTGAATTTGTCTCTCTCCATCCATCTCTCCATGCTGCTCACTTCTCTCCAGCTTC 1417  
Qy 1314 TTTGCTGCCCTATATCCATCTGAAATCCAACTGGAGCATCTACCCCAATACCA-- 1371  
Db 1418 TTTGGAGCGCATCACATACTGGACCTTCCCTCCAGGGGGAATCTACCCCAACCCCAAC 1477  
Qy 1372 -----GGCTGCCAGCTGCTCATATGCTTCCCATCTTGGACCTTACTACTA 1417  
Db 1478 GTCCCCCGCATCTTAACACCCAGTGTCTTACACITTAGGCACTTACTACTA 1530

RESULT 12  
US-10-342-887-710  
; Sequence 710, Application US/10342887  
; Publication No. US20040059340A1  
; GENERAL INFORMATION:  
; APPLICANT: Dai, Hongyue  
; APPLICANT: He, Yudong  
; APPLICANT: Linsley, Peter S.  
; APPLICANT: Mac, Mao  
; APPLICANT: Roberts, Christopher J.  
; APPLICANT: Van 't Veer, Laura Johanna  
; APPLICANT: Van de Vijver, Marc J.  
; APPLICANT: Bernards, Rene  
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients  
; FILE REFERENCE: 9301-188-999  
; CURRENT APPLICATION NUMBER: US/10/342,887  
; CURRENT FILING DATE: 2003-01-15  
; PRIOR APPLICATION NUMBER: 60/298,918  
; PRIOR FILING DATE: 2001-06-18  
; PRIOR APPLICATION NUMBER: 60/380,710  
; PRIOR FILING DATE: 2002-05-14  
; PRIOR APPLICATION NUMBER: 10/172,118  
; PRIOR FILING DATE: 2002-06-14  
; NUMBER OF SEQ ID NOS: 2699  
; SEQ ID NO 710  
; LENGTH: 2957  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-342-887-710

Query Match 38.9%; Score 562.2; DB 13; Length 2957;  
Best Local Similarity 66.4%; Pred. No. 4.9e-175;  
Matches 912; Conservative 0; Mismatches 428; Indels 33; Gaps 6;  
Qy 63 ATGCAAGCACTATTAAAGAGCATATCATGTGTGAGTGAAGACCACTGCTTTTGTAG 122  
Db 173 ATGACGGGACTATTAAAGAGGCTCTGTCTGGTGTGAGCGACGACCACTCTTTCAC 232  
Qy 123 TGTGCTACGGATCG- - -CCCACTTGAAGACAGAAATGACAGCTCTCTTCCAGT 179  
Db 233 TCAGGTACGGAGCGGAGCCCATCTCCCAAGCGCGACATGATGCTCGGGAGTCT 292  
Qy 180 GAATATGGGCAACATCAAGATGAGCCCGCGCTTCCCAAGAGGACTGTATATCAG 239  
Db 293 GACTACGGGCGGCCCAACAGATCAACCCCTCCCAACACAGCAGGAGTGGATCAATCAG 352  
Qy 240 CCCCCGGCAGAGTTACATTAAGTGGAGTGTACCCCAACACAGGTTAATGGTCAAG 299  
Db 353 CC-----AGTGAGGGTCAACGCTCAAGCGGGAGTATGCCACATGAATGGATCCAGG 403  
Qy 300 AATTCACTGTGACTGACGAGCGTGGCAAAAGAGGGAATAGTTAGCAGTTTCAGACAAT 359  
Db 404 GAGTCTCGGTGGAGTGCAGCGTTAGCAATAGCAGAGCTGTGGGCGGAGGCTCC 463  
Qy 360 GTTGGGATGAATATGGAAGTGTACATGGAAGAGAGCAT- - -ATTCCGCTTCAAAATG 416  
Db 464 AACCCCATGAATCAACAGCTATATGACGAGAGAAATGGCCCCCTCTCTCCCAACATG 523

RESULT 13  
US-10-172-118-710  
; Sequence 710, Application US/10172118

Qy 417 ACACCCATGAACGAGAGTATTGTGCCAGAGATCTTACGTTATGGAGCAGACACCAT 476  
Db 524 ACCACCAACGAGAGAGAGTATCTGTCCCGCAGACCCACACTGTGGACACAGAGCAT 583  
Qy 477 GTACGGCACTGCTGAGTGGGCACTGAAGAGTATGCTTCCAGACGCTGGACATCTTG 536  
Db 584 GTGAGCATGCTGGATGGGCCATTAAGAGTACAGCTTGATGGAGATCCACATCC 643  
Qy 537 TTGTTCCAGAAATTTGATGGGAAAGTTGTGTAAATGACCAAGATGACTTCCAGAGA 596  
Db 644 TTTTTCAGAACATGATGGCAAGAACTGTGTAAATGAAACAGAGGAGACTTCTCCGC 703  
Qy 597 CTCAGCCGAGCTATAACGACATATCTCTCTGTGTCACACTACACTACCTCAGAGAGA 656  
Db 704 GCCACACCCCTCTACACACGGAAGTGTGTGTCTCACCTCAGTTACTCTCAGGNAAG- 762  
Qy 657 GGAGCCACTTTTATTTTCCAAATACATCAGTTTACCAGAGAGCAACGCAAGATAACA 716  
Db 763 -----TTCACTGCTGGCTTATTAACACTCCACACCGCAATCTCTCAGATTGAGT 817  
Qy 717 ACAAGCCAGATTACCTTTATGAGCAAGAGATCAGCGTGGACAGTCCAGGCCAT 776  
Db 818 GTCAAGAAAGACCTCTTTATGACTCAGTCAGAGAGGAGCTTGGGCAATACATGAT 877  
Qy 777 CC- - -CACTCAGTCAAAAGCTACCCAAACCATCATCTTCAACAGTGCCTCAAAACAGAGAC 833  
Db 878 TCTGSCCTCAACAAAGTCTCTCCCTTGGAGGGGCAACCATCAGTAAGAATACAGAG 937  
Qy 834 CAGGCTCTCAGTTAGATCTTATCAGATTCTTGAACCGACAGCAGCGCTTTCGAAT 893  
Db 938 CAACGCCCCCAGCAGATCCGTATCAGATCTCTGGGCGGACAGCAGTCCGTAGGCAAC 997  
Qy 894 CCAGGAGTGGGCAATACAGTATGGCAGTTCTCTACTGGAGCTTCTGTGCGACAGCTCC 953  
Db 998 CTTGGAAGCGGCGAGATCCAGCTGTGGCAATCTCTCTGGAGCTGCTCTCCGACGCGC 1057  
Qy 954 AACTCCAACTGATCACTCTGGAGGGCAAAATGGGAGTTCAAGATGACAGACCTGAT 1013  
Db 1058 AAGCCAGCTGTATCACTCTGGAGGGGCAACCGGGAGTTCAAAATGACGAGCCCGAT 1117  
Qy 1014 GAAGTGGCTCGGCTTGGGAGAGAGGAAAGCAACCTAAACATGAACCTATGACAACTC 1073  
Db 1118 GAGTGGCCAGGCGCTGGGCGGCGGAGGAGCAACCAATGAATACGACAGCTG 1177  
Qy 1074 AGCGTGCATCTGCTTACTATGACAAAAATATATGACTAAAGTTTATGGTAAACGC 1133  
Db 1178 AGCGGGGCTCTCGT-TATTACTATGATAAAACATTTATGACCAAGTGCACGGCAAGA 1237  
Qy 1134 TATGCTTACAAATTTGATTTCCACGGAATCGCTCAGGCGCTCCAGCTCACCCTCCAGAA 1193  
Db 1238 TATGCTTACAAATTTGATTTCCACGCAATGCCAGCTTCCAGGCTCTCGAGCCATCGG 1297  
Qy 1194 TCATCCATGTACAAATPACCCATCAGACCTCCCTATACATGAGTTCTTACATGACACCC 1253  
Db 1298 TGTGTCATGTACAAGTACCTTCTGACATCTCTCATATGCTTCTTACCATGCCACAG 1357  
Qy 1254 CAGAAGATGAATTTGTAGTCCCATCCCTCTGCTTTGCGCGTAACTCATCCAGCTTT 1313  
Db 1358 CAGAAGTGAATTTGTCTCTCCCATCTCATCTCTCATGCTGTCTTCTCTCCAGCTTC 1417  
Qy 1314 TTTGCTGCCCTTAATCCATCTGGAATTCACCAACTGGAGGCACTTACCCCAATACCA- 1371  
Db 1418 TTTGAGCGGCATCAATACACTGGACCTCCCAACCGGGGGAATCTACCCCAACCCCAAC 1477  
Qy 1372 -----GGCTGCGAGCTGCTCATATGCTTCCCATCTTGGGCACTTACTACTA 1417  
Db 1478 GTCCCCCGCATCTTAACACCCAGTGTCTTACACTTAGGCACTTACTACTA 1530

```

; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172.118
; CURRENT FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/380, 770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 710
; LENGTH: 2957
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM 002017
; DATABASE ENTRY DATE: 2001-06-18
; US-10-172-118-710

Query Match 38.9%; Score 562.2; DB 13; Length 2957;
Best Local Similarity 66.4%; Pred. No. 4.9e-175;
Matches 912; Conservative 0; Mismatches 428; Indels 33; Gaps 6;

QY 63 ATGGCAACACTATTAAAGAGCATTATCAGTGTGAGTGAAGACCACTGCTGTTGAG 122
DB 173 ATGGACGGGACTATTAAAGAGCTCTGCGTGTGAGCGACCACTGCTGTTGAG 232
QY 123 TGTGCTACGAGTCG---CCCCACCTTGCAAGACAGAAATGACAGCTCTCTTCCAGT 179
DB 233 TCAGCGTACGGAGCGGCGCCATCTCCCAAGCGCGACATGATGCTCGTGGGAGTCT 292
QY 180 GAATATGGCAACATCAAGATGACCGGCGCTTCCCGAGCGAGTGTATATACAG 239
DB 293 GACTACGGGCGAGCCCAAGATCAACCCCTCCACACAGCGAGGATGATCAATCAG 352
QY 240 CCCCCGGCCAGAGTTACCAATTAAGATGAGTGTAAACCAACAGAGTTAATGGGTCAAG 299
DB 353 CC-----AGTGAGGCTCAACGTCAAGCGGGATGATGACCATGATGATCCAGG 403
QY 300 AATTCACTGATGACTGCGAGCGTGCAGCAAGAGGAGGAAATGGTTAGAGTTGACAAAT 359
DB 404 GAGTCTCCGGTGGACTGCGAGCGTTAGCAAAATGCAAGAGCTGGTGGCGGAGGCGAGTCC 463
QY 360 GTTGGATGAACATATGGAAGCTACATGGAAGAGAGCAT---ATTCCGCTCCCAATATG 416
DB 464 AACCCATGNACTACACAGCTATATGACGAGAGATGGCCCCCTCTCCCAACATG 523
QY 417 ACAACCAATGAACGAAGATGTTATTGTCGAGCAGATCTTACGTTATGAGCAGACAT 476
DB 524 ACCACCAACGAGAGGAGTATCTGTCGCGCAGACCCACACTGTGGACAGAGAGCAT 583
QY 477 GTACGCGAGTGGCTGAGTGGGCTGAGGATGATGTTCTCCAGAGTGGACATCTTG 536
DB 584 GTGAGGCAATGCTGGATGGGCCATAAGAGATACAGCTTGATGGAGATCGACATCC 643
QY 537 TTGTTCCAGAACATTGATGGGAAGAGTTGTGTAATATGACCAAGATGATTTCCAGAG 596
DB 644 TTTTCCAGAACATGATGGCAAGAACTGTGTAAATGAACAAGGAGGACTTCTCCGC 703
QY 597 CTCAGCGCGAGTATAAGCAGATATCTCTGTGTCACCTACACTACCTCAGAGAGAG 656
DB 704 GCCACCAACCTTACAAACGGAAGTGTGTGTGTCACACCTCAGTTACCTCAGGGAAG- 762
QY 657 GGAGCCACTTTTATTTTCCAAATACATCAGTTTACCAGAGCAAGCAACCAAGAAATACA 716
DB 763 -----TTCACTGCTGCTATATATACACCTCCACACCGCAATCTCTCAGATTGAGT 817

```

```

717 ACAAGCCAGATTTACCTTATGAGCAAGCGAGGAGATCAGCGTGCAGAGTGCAGCCAT 776
818 GTCAAGAAGACCCCTTTTATGACTCAGTCAAGAGGAGGCTTGGGCAATACATGAT 877
777 CC---CACTCAGTCAAAAGCTACCCCAACCATCATCTTCAACAGATGCCAAAACAGAGAC 833
878 TCTGGCTCAACAAAGATCTCTCCCTTGGAGGGGACAAACGATCAGTAAGATACAGAG 937
834 CAGCGTCTCAGTTAGTCTTATCAGATTCTTGGACGAGCAGCAGCAGCGTCTTCCAAAT 893
938 CACGCGCCCGCAGCAGATCCGATCAGATCTTGGGCGCCGACAGCAGTGCCTTAGCCAA 997
894 CCAGGAGTGGCAGATACAGCTATGGCAGTTCTTACTGGAGCTTCTGTGCGAGAGTCC 953
998 CTTGGAAGCGGCGAGATCCAGCTGTGGCAATCTCTCTGGAGCTGCTCTCCGACAGCGC 1057
954 AACTCCAACTGCATCACCTGGGAGGCGCAAAATGGGAGTTTCAAGATCAGACACCTCAT 1013
1058 AACGCCAGCTGTATCACCTGGGAGGGGACCAACGGGGAGTTCAAAATGCGACCCCGAT 1117
1014 GAAGTGGCTCGGCGTGGGAGAGAGAAACCAACCTTAAACATGAACCTATGACAAACTC 1073
1118 GAGGTGGCCAGCGCTGGGCGAGCGGAAACCAAGCCCAACATGAATTTACGACAGCTG 1177
1074 AGCGTGCATCTCGCTACTACTATGACAAAATATATGACTTAAAGTTTCATGGTAAACGC 1133
1178 AGCGGCGCTCTCGTTTACTTATGATATAAAACATTTATGACCAAAAGTGCACGGCAAGA 1237
1134 TATGCTCAAAATTTGATTTCCACGGAATCGCTCAGGCGCTCCAGCCTCACCTTCCAGAA 1193
1238 TATGCTTCAAAATTTGACTTCCACGCGATTTCCAGGCTCTGCGAGCCATCGGACCGAG 1297
1194 TCATCATGTAACAATACCCATCAGACTTCCCTCATGAGTTCCTACCATGACACCCCC 1253
1298 TCGTCCATGACAGTACCTTCTGACATCTCTCATGCTTCTTACCATGCCACCGAG 1357
1254 CAGAGATGAATTTGTAGTCTCCCATCCCTGCTTTGCGCGTAACTTACCTCAGCTTT 1313
1358 CAGAGTGAATTTGTCTCTCCATCCATCTCTCATGCTTCTTCTCTCTCAGCTTC 1417
1314 TTTGCTGCCCTTAATCCATCTGGAATTCACAACTGGAGGATCTTACCCCAATACCA-- 1371
1418 TTTGGAGCGCATCAATATCTGGAATCTCCCAACGGGGGGAATCTACCCCAACCCCAAC 1477
1372 -----GGCTGCCAGCTGCTCATATGCTTCCCTTCTTGGCAGCTTACTACTA 1417
1478 GTCCCCCGCCATCTTAACACCCAGTGCCTTTCACACTTAGGAGCTACTACTA 1530

RESULT 14
US-10-007-926A-295
; Sequence 295, Application US/10007926A
; Publication No. US20030143539A1
; GENERAL INFORMATION:
; APPLICANT: BERTUCCI, FRANCOIS
; APPLICANT: HOULGATTE, REMI
; APPLICANT: BIRNBAUM, DANIEL
; APPLICANT: NGUYEN, CATHERINE
; APPLICANT: VIENS, PATRICE
; APPLICANT: FEET, VINCENT
; TITLE OF INVENTION: GENE EXPRESSION PROFILING OF PRIMARY BREAST CARCINOMAS
; FILE REFERENCE: 1546-R-00
; CURRENT APPLICATION NUMBER: US/10/007,926A
; CURRENT FILING DATE: 2001-12-07
; PRIOR APPLICATION NUMBER: 60/254,090
; PRIOR FILING DATE: 2000-12-08
; NUMBER OF SEQ ID NOS: 468
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 295
; LENGTH: 2957
; TYPE: DNA

```



```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 18410
; LENGTH: 567
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP000163.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HEAT, SIGNAL = 0.92
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN B474, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.95
; OTHER INFORMATION: SWISSPROT HIT: P11308, EVALUE 9.00e-99
; OTHER INFORMATION: EST HUMAN HIT: R87572.1, EVALUE 0.00e+00
; OTHER INFORMATION: NT HIT: M17254.1, EVALUE 0.00e+00
US-09-864-761-18410

```

```

Query Match          26.7%; Score 387; DB 9; Length 567;
Best Local Similarity 83.7%; Pred. No. 2.9e-117;
Matches 438; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 896 AGGAGTGGGAGATACAGCTATGCGAGTTCCTACTGAGGCTTCTGCGAGACGCTCCAA 955
DB |||||
QY 523 AGGAGTGGGAGATCCAGCTTGGAGTTCTCTCTGAGGCTCTGTCGAGACGCTCCAA 464
DB |||||
QY 956 CTCGAATGCACTACCTGGGAGGACAAATGGGGAGTTCAAGATGACAGACCTGATGA 1015
DB |||||
QY 463 CTCGAATGCACTACCTGGGAGGACCAACCGGGAGTTCAAGATGACGGATCCCGACGA 404
DB |||||
QY 1016 AGTGCTGGGCTTGGGAGAGAGAAAGCAACCTAACTGAACTATGACAAACTCAG 1075
DB |||||
QY 403 GGTGCGCGCGCTGGGAGAGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 344
DB |||||
QY 1076 CCGTCACTTCGCTACTACTATGACAAATAATTATGACTAAAGTTATGTTAAACGCTA 1135
DB |||||
QY 343 CCGCGCCCTCCGTTACTACTATGACAGAGACATCATGACCAAGGTCCATGGGAAGCGCTA 284
DB |||||
QY 1136 TGCCTACAAATTTGATTTCCAGGAGTCCGAGGCTCCAGGCTCCAGCTCCAGGATC 1195
DB |||||
QY 283 CGCCTACAAAGTTCGACTTCCAGGGATCGCCAGGCTCCAGGCTCCAGGCTCCAGGCTC 224
DB |||||
QY 1196 ATCCATGTACAAATACCCATCAGACCTCCCTTACATGAGTTCTTACCATGACACCCCA 1255
DB |||||
QY 1256 GAAGATGAACTTGTAGTCCGCTCCGCTCCGCTTTCGCGTACCTCATCCAGCTTTT 1315
DB |||||
QY 163 GAAGATGAACTTGTGGGCGCCCGCCAGCTCCAGGCTCCCGGAGATCTCCAGTTT 104
DB |||||
QY 1316 TGTGCCCCATATCCATACCTAGGAAATTCACCACTGGAGGATCTACCCCAATACAGGCT 1375
DB |||||
QY 103 TGTGCCCCAAACCCATACCTAGGAAATTCACCACTGGGATATATACCCCAACACTAGGCT 44
DB |||||
QY 1376 GCCAGTGTCTATATGCTTCCCTTCCAGCTTGGCACTACTACTAA 1418
DB |||||
QY 43 CCCACAGCCATATGCCTTCTCATCTGGGCACTTACTACTAA 1
DB |||||

```

Search completed: May 27, 2004, 21:11:29  
Job time : 732 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 27, 2004, 16:28:06 ; Search time 4096 Seconds  
(without alignments)  
10549.457 Million cell updates/sec

Title: US-09-902-772-1

Perfect score: 1447

Sequence: 1 gaattccgcgaagcaataat.....gaaagaagcgccaagaaaa 1447

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_hic:\*

9: gb\_estl:\*

10: gb\_est2:\*

11: gb\_hic:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_ges\_hum:\*

18: em\_ges\_inv:\*

19: em\_ges\_pln:\*

20: em\_ges\_vrt:\*

21: em\_ges\_fun:\*

22: em\_ges\_mam:\*

23: em\_ges\_mus:\*

24: em\_ges\_pro:\*

25: em\_ges\_rod:\*

26: em\_ges\_phg:\*

27: em\_ges\_vrl:\*

28: gb\_gesl:\*

29: gb\_ges2:\*

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	917.2	63.4	1350	29	AY419454 Homo sapi
2	881.6	60.9	1866	11	AK050922 Mus muscu
3	879.4	60.8	1350	29	AY419456
4	876.4	60.6	3064	11	AK076113 Mus muscu

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

5	856	59.2	1350	29	AY419455
6	692.4	47.9	726	13	BU334592
7	570.2	39.4	1706	11	AK042111
8	560.6	38.7	1347	29	AY418895
9	559	38.6	3118	11	AK036655
10	552.6	38.2	1347	29	AY418893
11	526.2	36.4	606	13	BU478556
12	501.2	34.6	1256	29	AY418894
13	485.4	33.5	980	13	BQ233264
14	483.4	33.4	709	13	CE265338
15	482.2	33.3	715	12	BG389025
16	475.4	32.9	642	12	BT13036
17	474.4	32.8	521	13	BU440626
18	468.6	32.4	701	9	AU136709
19	468.6	32.4	1089	13	BQ212127
20	462	31.9	634	12	BM489636
21	457.4	31.6	810	13	BU330912
22	452.2	31.3	668	10	BE660034
23	445.2	30.8	915	12	BG390291
24	444.2	30.7	928	12	BG259459
25	442	30.5	1045	12	BM456833
26	438.4	30.3	877	13	BU332566
27	437.4	30.2	567	13	BU336313
28	436.2	30.1	675	10	BG639043
29	432.6	29.9	629	10	BE577887
30	428.8	29.6	629	10	AW948986
31	425.6	29.4	592	12	BM088063
32	403.4	27.9	650	13	BU490135
33	400.2	27.7	616	14	CB579500
34	393.6	27.2	602	14	CB582161
35	389.8	26.9	2420	11	AK083900
36	385.8	26.7	526	14	CA405791
37	384.4	26.6	802	14	CB952600
38	380	26.3	827	9	AJ456498
39	367.8	25.4	522	10	BF566705
40	367.2	25.4	516	12	BG691566
41	361	24.9	693	13	BX100094
42	359.6	24.9	560	9	AU280022
43	358.8	24.8	976	13	BX328036
44	354	24.5	1406	28	CC264217
45	352.8	24.4	949	12	BI558012

#### ALIGNMENTS

RESULT 1	AY419454	1350 bp	DNA	linear	GSS 17-DEC-2003
LOCUS	AY419454				
DEFINITION	Homo sapiens ERG gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.				
ACCESSION	AY419454				
VERSION	AY419454.1				GI:39775411
KEYWORDS	GSS.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Bukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejariwal, A., Todd, M.A., Tanenbaum, D.M., Civallo, D.R., Lu, F., Murphy, B., Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M.				
TITLE	Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios				
JOURNAL	Science 302 (5652), 1960-1963 (2003)				
PUBMED	14671302				
REFERENCE	2 (bases 1 to 1350)				
AUTHORS	Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejariwal, A., Todd, M.A., Tanenbaum, D.M., Civallo, D.R., Lu, F., Murphy, B., Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M.				
TITLE	Direct Submission				

JOURNAL	Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA									
COMMENT	These sequences were made by sequencing genomic exons and ordering them based on alignment.									
FEATURES	Location/Qualifiers									
	1..1350									
source	/organism="Homo sapiens"									
	/mol_type="genomic DNA"									
gene	/db_xref="taxon:9606"									
	<1..>1350									
ORIGIN	/gene="ERG"									
	/locus_tag="HCM6996"									
Query Match	63.4%; Score 917.2; DB 29; Length 1350;									
Best Local Similarity	79.1%; Pred. No. 3.5e-263;									
Matches 1668; Conservative	0; Mismatches 270; Indels 12; Gaps 2;									
QY	81	GAAGCATTATCAGTGGTGAAGACCAAGCCAGTCCCTGTTGTTGAGTGTGCTACGGATCCGCC	140							
Db	1	GAAGCCTTATCAGTTGTGAGTGAGGACCAAGTCCGTTGTTGAGTGTGCTACGGAAACGCCA	60							
QY	141	CACCTTGCAAGACAGAAATGACAGCCCTCTTCCAGTGAATATGGCAACATCAAG	200							
Db	61	CACCTGGCTAAGACAGATGACCGGCTCTCTCCAGGACTATGGACACTTCCAAAG	120							
QY	201	ATGAGCCGCGGCTTCCCGCAGGAGCTGTTATCACAGCCCGCGGCGAGATTACCAAT	260							
Db	121	ATGAGCCCGAGCGGCTCCCTCAGCAGATTGGTGTCTCAACCCCGCAGCGAGGTCAACATC	180							
QY	261	AGATGGAGTGTAAACCCAAACAGGTTAATGGGTCAAGGAATTCACCTGATGATGGCAGC	320							
Db	181	AAATGGAAATGTAACTTACCTAGCAGGTGAATGGCTCAAGGAACTCTCTGTGATGAATGAGT	240							
QY	321	GTGGCAAAAGAGGAGAAATGGTTAGCAGTTTACAGCAATGTTGGGATGAATATGGGAAGC	380							
Db	241	GTGGCCAAAGGCGGGAAGATGGTGGGAGCCAGACACCCGTGGGATGAATCAAGCAGC	300							
QY	381	TATCGTGAAGAGAGCATATTCGCTCCCAATATGACAAACATGAAGAAAGAGTTATT	440							
Db	301	TATCGGAGGAGAGACATGACACCCCAACATGACACGAAACGAGCGCAGAGTTATC	360							
QY	441	GTGCGAGCAGATCTTACGTTATGAGAGACAGACCATGTACGCGAGTGTGCTGGATGGGCA	500							
Db	361	GTGCGCAGCATCTTACGCTATGAGTACAGACCATGTGCGGCAATGTGGGAGTGGGCG	420							
QY	501	GTGAAGGAGTATGCTTCCAGACGTGGACATCTTTGTTGTTCAGAAATGTATGGGAA	560							
Db	421	GTGAAGAATATGGCTTCCAGACGTCAACATCTTTGTTATTCAGAAACATCGATGGGAAG	480							
QY	561	GAGTTGTGTAATGACCAAGATGACTTCCAGAGACTCACGCGAGCTATAACGCGAGAT	620							
Db	481	GAACTGTGAAGATGACCAAGAGCACTTCCAGAGCTCACCCAGCTACACGCGGAC	540							
QY	621	ATCCTCTCTGTACACCTACACTACCTCAGAGAGA-----GAGGAGCCACTTTTATT	671							
Db	541	ATCCTTCTCTCATCTCCACTACCTCAGAGAGANNNNNNNNNNNNNNNNNNNNNNNN	600							
QY	671	TTTCCAAATACATCAGTTTACCCAGACGACGCAAGCAATTAACCAAGCCAGATTTA	731							
Db	601	NN	660							
QY	731	CCTTATGAGCAAGCGAGGAGATCAGCGTGGAGAGTACAGCCATCCCA---CTCAGTCA	788							
Db	661	CCATATGAGCCCGCCAGGAGATCAGCTGGACCGGTACGCGCCACCCCGCCCGAGTCG	720							
QY	788	AAAGCTACCAACCATCATCTTCAACAGTGGCCCAAAACAGAGACCGGCTCTCAGTTA	848							
Db	721	AAAGCTGTCTCAACCATCTCTCTTCCAGTGGCCCAAACTGAAGACCGGCTCTCAGTTA	780							
QY	848	GATCCTTATCAGATCTTGGACCGGACAGCAGCGCTTGTGAAATCCAGGAGTGGGCGAG	908							
Db	781	GATCCTTATCAGATCTTGGACCAACAGTAGTACCGGCTTGCATCTCAGGAGTGGCGAG	840							

QY 909 ATACAGCTATGCGAGTCTCTACTGAGCTTCTGTGCGACAGCTCCAACTCCAACTGCATC 968  
Db 841 ATCCAGCTTTGGAGTCTCTCTGAGCTTCTGTGCGACAGCTCCAACTCCAACTGCATC 900  
QY 969 ACCTGGAGGCGACAAATGGGAGTTCAAGATGACAGACCCCTGATGAAGTGTGCTGGGCT 1028  
Db 901 ACCTGGAGGCGACAAACGGGAGTTCAAGATGACAGGATCCCGAGAGTGGCGCGCGC 960  
QY 1029 TGGGAGAGAGAAAGCAACCTAACATGAACTGACAACTCAGCCGTCACCTTGC 1088  
Db 961 TGGGAGAGCGGAAGAGCAACCCAACTGAACTAGATAGCTCAGCGCGCCCTCGT 1020  
QY 1089 TACTACTATGACAAATAATTATGACTAAAGTTTCAATGTTTCAATGCTATGCTTACAAATTT 1148  
Db 1021 TACTACTATGACAAAGACATCATGACCAAGTTCCTATGGAAGCGCTACGCTTACAAATTC 1080  
QY 1149 GATTTCCAGGAATGCTCAGCCCTCCAGCTCCAGCTCCAGCTCCAGCTCCAGCTCCAGCT 1208  
Db 1081 GATTTCCAGGAATGCTCAGCCCTCCAGCTCCAGCTCCAGCTCCAGCTCCAGCTCCAGCT 1140  
QY 1209 TACCCATCAGACTCCCTTACATGATGCTTCTTACCATGACACCCCGCAGAGATGAATCTTT 1268  
Db 1141 TACCCCTCAGACTCCCTTACATGATGCTTCTTACCATGACACCCCGCAGAGATGAATCTTT 1200  
QY 1269 GTAGCTTCCCATCTCCCTTGTGCGGTAACCTCATCCAGCTTTTTCCTGCTGCTTAT 1328  
Db 1201 GTGCGCGCCCGCTCCAGCTCCAGCTCCAGCTCCAGCTTTTTCCTGCTGCTTAT 1260  
QY 1329 CCATCTGGAATTCACCAACTGGAGGATCTTACCCCAATACCAGGCTGCGCTCAT 1388  
Db 1261 CCATCTGGAATTCACCAACTGGGATATATACCCCACTAGGCTCCCGCAGCCAT 1320  
QY 1389 ATGCTTCCCATCTTGGCAGCTTACTACTAA 1418  
Db 1321 ATGCTTCTCATCTGGGCACTTACTACTAA 1350

RESULT 2  
AK050922 LOCUS  
DEFINITION  
Mus musculus 9 days embryo whole body cDNA, RIKEN full-length  
enriched library, clone:D030036124 product:ERG mRNA, mouse homolog  
of Human ets-related gene ERG, transcript variant 1, full insert  
sequence.  
AK050922  
AK050922.1 GI:26341597  
HTC; CAP trapper.  
Mus musculus (house mouse)  
MUS MUSCULUS  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
99279253  
MEDLINE  
10349636  
PUBMED  
REFERENCE  
2  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
20499374  
MEDLINE  
11042159  
PUBMED  
REFERENCE  
3  
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Kusunagi, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamanoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanishi, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,



Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)  
 20530913  
 MEDLINE  
 PUBMED  
 11076861  
 4  
 REFERENCE  
 AUTHORS

The RIKEN Genome Exploration Research Group Phase II Team and the  
 FANTOM Consortium.  
 Functional annotation of a full-length mouse cDNA collection  
 Nature 409, 685-690 (2001)  
 5  
 REFERENCE  
 AUTHORS

The FANTOM Consortium and the RIKEN Genome Exploration Research  
 Group Phase I & II Team.  
 Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)  
 6 (bases 1 to 1866)  
 ADACHI, J., AIZAWA, K., AKIMURA, T., ATAKAWA, T., BONO, H., CARNINCI, P.,  
 FUKUDA, S., FURUNO, M., HANAGAKI, T., HARA, A., HASHIZUME, W.,  
 HAYASHIDA, K., HAYATSU, N., HIRAMOTO, K., HIRAOKA, T., HIROZANE, T.,  
 HORI, F., IMOTANI, K., ISHII, Y., ITOH, M., KAGAWA, I., KASUKAWA, I.,  
 KATO, H., KAWAI, J., KOJIMA, Y., KONDO, S., KONNO, H., KODA, M.,  
 KOYA, S., KURIHARA, C., MATSUYAMA, T., MIYAZAKI, A., MURATA, M.,  
 NAKAMURA, M., NISHI, K., NOMURA, K., NUMAZAKI, R., OHNO, M., OHSATO, N.,  
 OKAZAKI, Y., SAITO, R., SAITO, H., SAITO, H., SAKAI, C., SAKAI, K., SAKAZUME, N.,  
 SANO, H., SASAKI, D., SHIBATA, K., SHINAGAWA, A., SHIRAKI, T.,  
 SOGABE, Y., TAGAMI, M., TAGAWA, A., TAKAHASHI, F., TAKAKU-AKAHIRA, S.,  
 TAKEDA, Y., TANAKA, T., TOMARU, A., TOYA, I., YASUNISHI, A.,  
 MURAMATSU, M. and HAYASHIZAKI, Y.  
 Direct Submission

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
 Physical and Chemical Research (RIKEN), Laboratory for Genome  
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
 Kanagawa 230-0045, Japan (E-mail: genome-res@gs.c.riken.go.jp,  
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,  
 Fax: 81-45-503-9216)  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.  
 Please visit our web site for further details.  
 URL: http://genome.gsc.riken.go.jp/  
 URL: http://phantom.gsc.riken.go.jp/  
 Location/Qualifiers

1..1866  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="FANTOM DB:D030036124"  
 /db\_xref="MGI:2418825"  
 /db\_xref="taxon:10090"  
 /clone="D030036124"  
 /tissue\_type="whole body"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="9 days embryo"  
 249..1637  
 /note="unnamed protein product; Erg mRNA, mouse homolog of  
 Human ets-related gene ERG, transcript variant 1  
 (DBJ|AB073078, evidence: BLASTN, 99%, match=1768)  
 putative"  
 /codon\_start=1  
 /protein\_id="BAC34461.1"  
 /translation="MIQVPPAAHIKEALSIVSDQSLFECAYGTPELAKTEMTASS  
 SSDYGTQSKMSPVQQDLWLSQAPARVTIKMECPNSOVNGSRNPDECSVNGKGMWV  
 SPDYVMSVGSVMKEXYKPPDNTNERRIVLPADPTLWSDHVMLEWAKVYGLL  
 DVDFLFQNDIGNEICKYKTKDQFRLTFSYNADILLSHLYLRSTPLPRLTSDVDKA  
 LQNSRLMHRNTDLPYEPFRSAWTHSHLTPQSKAAQSPSPAVPKTDEPOLDFY  
 QILGPTSRSLNPGSQIQWQFLLELLSDNSNCNTWETNGEFGKMDDEVARW  
 GERKSKPNMNYDKLSRLRYYYDKNIMTKVHKRYAYKFPDFHGAQALQPHPPSSSLY

Query Match 60.9%; Score 881.6; DB 11; Length 1866;  
 Best Local Similarity 79.3%; Pred. No. 2e-252;  
 Matches 1075; Conservative 0; Mismatches 269; Indels 12; Gaps 2;  
 KyPSPDLPMYGSYHAHPQKQNFVSPHPPLPVTSSFFSPNPYWNPSPTGGIYENTLPL  
 ASHMPSHLGTYY"

ORIGIN

Qy 74 TATTAAAGGAAGATTATCAGTGGTGAAGTGAAGACCAAGTCTCTTGTGTTGAGTGTGCTTACCG 133  
 Db 281 TATTAAAGGAGGCTTGTCTGAGTTGTGAGCGAGGACCACTATTTGAGTGTGCTTACCG 340

Qy 134 ATCGCCCAACCTTGTCAAGACAGAAATGACAGCTCTCTTCCAGTGAATATATGGGCAAAAC 193  
 Db 341 AACCCACACCTGGCTTAAGACAGAGATGACCGCATCTCTTCCAGTGAATATATGGCAGAC 400

Qy 194 ATCAAAAGATGAGCCCGCGTTCCTCCAGCAGGACTGGTATTCACAGCCCCCGGCGAGAGT 253  
 Db 401 ATCAAGATGAGTCCCGAGAGTCCCTCAGCAGGACTGGTGTCTCAAGCCCCCGAGGAGGT 460

Qy 254 TACCATTAGATGGAGTGTAAACCCAAACCCAGGTTAATGGTCAAGGAATTCACCTCATGA 313  
 Db 461 CACCATCAAGATGGAGTGTCAACCCCTAGTCAGGTGAATGGTTCAGGAATTCACCTCATGA 520

Qy 314 CTGCAGCGTGGCAAAAGAGGAGGAAATATGTTAGCAGTTTCCAGCAATGTTGGGATGAACCTA 373  
 Db 521 GTCCAGTGTGAACAAAGGTGGGAAGATGTTGGSCAGCCCGGATCTGTGGGATGAGCTA 580

Qy 374 TGGAACTCATGGAAGACAGAGCATATTCGGCTCCAAATATGACAAACCAATGAACGAAG 433  
 Db 581 CGCAGCTCATGGAAGAGAGCATGTGCGGCTCCCAATATGACCAACCAATGAGCGCAG 640

Qy 434 AGTTATTGTGCCAGCAGATCTTACGTTTATGGAGCAGACCATGTATCGGCAGTGGTGA 493  
 Db 641 AGTGATCGTCCCTGCAGATCTTACTCTGTGGAGCAGACCATGTCCGACAGTGGCTGA 700

Qy 494 GTGGGAGTGAAGGATGTGTTTCCAGACGTGGACATCTTGTGTTTCCAGAACATTGA 553  
 Db 701 GTGGGCGGTGAAGAATATGGCTCTCTCATGTGGACGCTTACTATTTCAGAAATATGA 760

Qy 554 TGGGAAGAGTGTGTGTAATAAGCAAGATGACTTCCAGAGACTCACGCCGAGCTATAA 613  
 Db 761 TGGGAAGAGCTGTGCAAGATGACAAAGATGACTTCCAGCGGCTCACGCCGAGCTACAA 820

Qy 614 CGCAGATATCTCTCTGTCAACCTACATCTCTCAGAGAGACTCCCTTCCACATCTGAC 664  
 Db 821 TGCCGACATCTTCTCTCATCTCCATCTCCACTACCTCAGAGAGACTCCCTTCCACATCTGAC 880

Qy 665 TTTTATTTTCCAAATACATCAGTTTACCAGAGCAAGCGCAAGAAATACACAAAGGCC 724  
 Db 881 TTCGATGACGTTGTAAGGCTTTACAAACTCTCCCGTAAATGATGCGCAGAAACAC 940

Qy 725 AGATTTACCTTATGAGCAAGCGAGGAGATCAGCGTGGAGAGTCAAGCAGATC---CCAC 781  
 Db 941 AGATTTACCTTATGAGCTCCAGGAGATCAGCGTGGAGAGTCAAGCAGATCAGGCTCC 1000

Qy 782 TCAGTCAAAAGCTACCCCAACCATCTTCAACAGTGGCCCAACCAAGAGACAGCGGTCC 841  
 Db 1001 TCAGTCAAAAGCTGCTCAGCCATCTCCCTCTGAGTGGCCCAACCAAGAGAGAGGCTCC 1060

Qy 842 TCAGTTAGATCTTATCAGATTCTTGGACCGACAGCAGCGCTTTCGAAATCCAGGAG 901  
 Db 1061 TCAGTTAGATCTTACCAGATCTGGGACCGACAGTAGCGCGCTTGTCTAATCCAGGTAG 1120

Qy 902 TGGCAGATACAGCTATGGCAGTCTTCTACTGGAGCTTCTGCGAGAGCTCCAACTCCAA 961  
 Db 1121 TGGCCAGATCCAGCTGTGGCAGTCTCTGCTCGCAACTCTCTGAGAGCTCCAACTCCAA 1180

Qy 962 CTGCATCACCTGGAGGGGCAAAATGGGAGTTCAGAGTACAGACAGCCCTCATGAAGTGGC 1021  
 Db 1181 CTGCATCACCTGGAGGGCACCAACGGGAGTTCAGATGACAGACAGCCGAGGAGTGGC 1240

Qy 1022 TCGGCGTTGGGAGAGAGGAAAAACCACTAATCATGAATATGACAACTCAGCCGTGC 1081

Db	1241	TCGGCGCTGGGGGAGAGGAGCAAGCCCAACATGAATGACCAAGCTCAGCGCGC	1300	81	GAAGCATTATCAGTGGTGAAGACCAAGTCCTTTGTTGAGTGTGCTACGGATCGCC	140
Qy	1082	ACTTCGTACTACTATGACAAATATATGACTTATGTTTCATGGTAAACGCTATGCTA	1141	1	GAGGCTTGTGCTAGTGTGAGGGAGGAGCCAGTCACTATTGTTGAGTGTGCTACGGAACGCA	60
Db	1301	CCTCGGTACTACTACGACAAACATCATGACCAAGTGCAGGAGGCTACGCTA	1360	141	CACCTTGCACAGACAGAAATGACAGCCTCTCTTCCAGTGAATATGGGCAACATCAAG	200
Qy	1142	CAAAATTTGATTTCACAGGAATCGTCAAGGCTCCAGCCTCACCTCCAGAAATCATCCAT	1201	61	CACCTGGCTAAGACAGAGATGACCGCATCTCTTCCAGTGAATATGGGCAACATCAAG	120
Db	1361	CAAGTTTGACTTCACAGGATTCAGGCGCTGCGCCAGCCACCTCTGAGTGTGCT	1420	201	ATGAGCCCGCGGTTCCTCCAGCAGGAGTGTATACAGCCCCCGGCGCAGAGTTACCAT	260
Qy	1202	GTACAAATACCATCAGACCTCCCTACATGAGTTCCTACATGACACACCCAGAGAT	1261	121	ATGAGTCCAGAGTCCCTCAGCAGGAGTGTGCTCAAGCCCGCAGCCAGGCTCACCAT	180
Db	1421	GTACAGTACCTCCAGCCTGCTCATATGAGTGTGCTCATGAGTGTGCTCATGAGT	1480	261	AGATGGAGTGTAAACCCAAACAGGTTAATGGTCAAGGAATTCACCTGATGAGTGCAGC	320
Qy	1262	GAACTTTGTAGTCCCATCCCTGCTTTGCCGTAACTCATCCAGCTTTTGTGCTC	1321	181	AGATGGAGTGTCAACCTTAGTGTGAGTGAATGGTCCAGAACTCACCTGATGAGTGCAGT	240
Db	1481	GAACTTTGTGCTCCACACCTCCCGCTCTCCAGTCAATCTTCCAGTTTCTTGTGCTC	1540	321	GTGCAAAAGGAGGAAATGGTTAGCAGTTTCAAGCAATGTGGGATGAATATGGAAGC	380
Qy	1322	CCCTAATCCATCTGGAATTCACCACTGGAGGCTATCCCAATACCAAGCTGCGCAGC	1381	241	GTGCAAAAGGTTGGGAGATGGTGGGAGCGCGGATCTGTGGGATGAGTGTACGCGCAGC	300
Db	1541	CCGGAACCATCTGGAATTCACCACTGGAGGCTATCCCAATACCAAGCTGCGCAGC	1600	381	TACATGGAGAGAGCATATTCGCGCTCCAAATATGACAAACCAATGAACGAGAGTATT	440
Qy	1382	TGCTCATATGCTTCCCATCTTGGCAGCTACTACTA	1417	301	TACATGGAGAGAGCATATTCGCGCTCCAAATATGACAAACCAATGAACGAGAGTATT	360
Db	1601	CAGCCATATGCTTCCATCTCAGCTGGGACCTACTACTA	1636	441	GTGCCAGCAGATCTTACGTTATGGAGACAGACCAATGTACGCGAGTGGCTGGAGTGGCA	500
RESULT 3				361	GTCCCTGCAGATCTTCTGTGGAGACAGACCAATGTCCGAGTGTGGTGGGCG	420
AV419456				501	GTGAGGAGTATGGTCTTCCAGACGTGACATCTTGTGTTCCAGAAATGATGATGGGAA	560
LOCUS				421	GTGAAAGAAATATGGGCTCTCTCGATGTGAGCGTCTTACTATTTCAGAAATCATGATGGGAG	480
DEFINITION		1350 bp DNA linear	GSS 17-DEC-2003	561	GAGTTGTGTAAATGACCAAGATGACTTCCAGAGACTCACGCCGAGCTATTAACGAGAT	620
ACCESSION				481	GAGCTGTGAGATGACAAAGATGACTTCCAGCGCTCACGCCGAGCTACAATGCCGAG	540
VERSION				621	ATCCTCTGTGCACCTACATCTACCTCAGAGAGA-----GAGGAGCCACTTTTATT	571
KEYWORDS				541	ATTCTTCTCTCATCTCTCCACTACCTCAGAGAGANNNNNNNNNNNNNNNNNNNNNNNN	600
SOURCE				672	TTTCCAAATACATCAGTTTACCAGAGCAACGCAAGAAATAACAAAGGCCAGATTATA	731
ORGANISM				601	NN	660
REFERENCE				732	CCTTATGACAAAGCAGGAGATCAGCGTGGAGAGTCAAGCCATC---CCTCAGTCA	788
AUTHORS				661	CCTTATGAGCTTCCAGGAGATCAGCTGGAGCCGCGCCACAGCCACTCACCCCTCAGTCC	720
TITLE				789	AAAGCTACCCACCATCATCTTCAACAGTGGCCCAAAACAGAGACAGCGTCTCAGATT	848
JOURNAL				721	AAAGCTGTCTCAGCCTATCTCCTCTGAGTGGCCCAAACTGAAGACAGCGTCTCAGTTA	780
PUBMED				849	GATCCTTATCAGATTCTTGGACCGACAGCCGCTTTGCAAAATCCAGGAGTGGGCGAG	908
REFERENCE				781	GATCCTTATCAGATTCTTGGACCGACAGCCGCTTTGCAAAATCCAGGAGTGGGCGAG	840
AUTHORS				909	ATACAGTATGAGTGTCTCTTCTGAGTGTCTGAGAGTCTGAGTGTCTGAGTGTCTGAGT	968
TITLE				841	ATCAGCTGTGGAGTCTCTGCTCGAATCTCTGTCAGACAGTCTTCACTTCACTTCACT	900
JOURNAL				969	ACCTGGGAGGACCAAAATGGGAGTTCAGATGACAGACCTGATGAAGTGGCTCGGCT	1028
COMMENT				901	ACCTGGGAGGACCAAAATGGGAGTTCAGATGACAGACCTGATGAAGTGGCTCGGCT	960
FEATURES				1029	TGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1088
source				961	TGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1020
gene				1089	TACTACTATGACAAATATTTATGACTAAAGTTTCAATGTTTCAATGTTTCAATGTTT	1148
ORIGIN				1021	TACTACTACGACAAACATCATGACCAAGTGTGACGGAAGCGCTACGCTTACAGATT	1080
Query Match				1149	GATTTCCAGGAATCGCTCAGGCGCTCCAGCCTCCAGATCATCCATGATCAAA	1208
Best Local Similarity						
Matches 1044;						
Conservative						
0;						
Mismatches						
293;						
Indels						
12;						
Gaps						
2;						

AUTHORS	Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanaoka,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitho,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE	Direct Submission
JOURNAL	Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL:http://genome.gsc.riken.go.jp/ URL:http://fantom.gsc.riken.go.jp/ Location/Qualifiers 1. 3064 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM DB:6330501C04" /db_xref="MGI:239312" /db_xref="taxon:10090" /clone="6330501C04" /sex="male" /tissue_type="medulla oblongata" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="adult" 61. 1431 /note="unnamed protein product; Erg mRNA, mouse homolog of Human <i>erg</i> -related gene <i>ERG</i> , transcript variant 1 (DBJ AB073078, evidence: BLASTN, 99%, match=1768) putative" /codon_start=1 /protein_id="BAC37131.1" /db_xref="GI:26346965" /translation="WASPIKEALVSRSQSLFECAVGTPLAKTMTMTSSSSDYGOT SYGSRVFEQHWLSQAPARVITKMCNSQVNSRSPDECSVNGKVMGSDPTVGM KMSVTRVQKHWPPPPNNTNERRVITPFDPLMTSDHWROWLEWAWKEYGLLDVLLF INDIQEGKTKTDQFRLTSYNADILLSHLHIRELPPLHLTSDVDKALQNSRL MHARTGGAATFPNTVSPATQITTRPAAPQSPFAVPTEDQRPOLDPQVILGET SRRLNPGGIGIQLMVQLELLSDSNGCTIWEQTEGFKMTDPDEAVRWGERSK SRMLVDKLSRLAYVYDKNIMTKVHGKYAVYDFHGIQAALQHPPESSLYKPSDL PYMGSYHAHPQCMNFVSPHPALPTSSFFPASPYPNYSPTGGIYVNTRELPSHESMS HLGTYY"
FEATURES	
source	
CDS	
ORIGIN	
Query Match	60.6%; Score 876.4; DB 11; Length 3064;
Best Local Similarity	77.1%; Pred. No. 9.1e-251;
Matches 1101; Conservative	0; Mismatches 306; Indels 21; Gaps 2;
QY	8 GCGAAGCAATAATTATTATTCACAAATTATTAGCGATCAATAATCTTGATCACAATTATGCC 67
Db	6 GAGAAGCAATAATTATTATTCACAAATTATTAGCGATCAATAATCTTGATTCATTATGCC 65
QY	68 AAGCACTATTAAGGAAGCAATTATCAGTCGTGTGAGTGAGGACCAGTCCCTTGTTTCAGTGTGC 127
Db	66 CAGCACTATTAAAGAGGCCCTTGTGAGTCGTGTGAGCGAGGACCAGTCACTATTGAGTGTGC 125
QY	128 CTACGAGATCGCCCACTTTCGAAAGACAGAAATGACAGCCTCTCTTCCAGTGAATAATGG 187

126 CTACGGAAACCCACCTGGCTAAGACAGAGATGACCGCATCTCTTCCAGTACTATGG 185  
188 GCAAAATCAAGATGAGCCCGCGTTCGCCAGCAGGACTGTTATCACAGCCCCCGGC 247  
186 CCAGACATCCAAAGATGAGTCCAGAGTCCCTCAGCAGGACTGGCTGTCTCAAGCCCCAGC 245  
248 CAGAGTTACCATTAAGATGGAGTGAACCCAAACCCAGTTTAAATGGGTCAGAGTATCAAC 307  
246 CAGGTCACCATCAAGATGGAGTGAACCTTAGTCAGGTGAATGGTTCCAGGAACTCACC 305  
308 TGATGACTGACGCGTGGCAAAAGAGGAGAAATGTTAGCAGTTTCAGACAAATGTTGGAT 367  
306 TGATGAGTGCAGTGTGAACAAAGTGGAGATGGTGGCAGCCCGGATACTGTGGGAT 365  
368 GAATATGGAGCTACATGGAGAGAGCATATTCGCCCTCCAAATATGACACCAATGA 427  
366 GAGCTACGGCAGCTACATGGAGGAGAGCATGTGCCCGCTCCCAATATGACCAAAATGA 425  
428 ACGAAGAGTTATTGTCCAGCAGATCCTACGTTATGGAGCAGACAGACCATGTACGGCAGTG 487  
426 GCGCAGAGTATCTGCTCCCTGAGATCTTACTCTGTGGAGCAGACCATGTCCGACAGTG 485  
488 GCTGAGTGGCAGTGAAGAGTATGTTCTCCAGAGCTGAGACATCTTGTGTTTCAGAA 547  
486 GCTGAGTGGCGGTGAAGAAATATGGCTCTCTCGATGTGACGTCTTACTATTTCAGAA 545  
548 CATTGATGGAAAGAGTTGTCTAAATGACCAAGATGACTTCCAGAGACTTCAGCCGAG 607  
546 TATCATGGAGAGAGCAGTCAAGATGACAAAGATGACTTCCAGCGGCTCAGCCGAG 605  
608 CTATAACGAGATATCTCTCTGTCTCACACTACATCTACCTCAGAGAGAGAGCCACTTT 667  
606 CTACAATCCGACATCTCTCTCACATCTCCACTACCTCAGAGAGACTCCCTTCCACA 665  
668 TATTTTCCAAATACATCAGTTTACCCAGAGCAA-----CGCAAG 709  
666 TCTGACTTCCGATGACGTTGTAAGGTTTACAAATCTTCCAGGTTAATGATGCCAG 725  
710 AATAACAAACAGGCGAGATTTACCTTATGACAAAGCAGGAGATCAGCGTGGACGATCA 769  
726 AAACACAGGGGTGCAGCTTTTATTTTCCAAATACTTCAGTATATCCGAAGTACGCA 785  
770 CAGCCATCCACTCAGTCAAAAGTACCCACATCATCTTCAACAGTCCCAACAGAGA 829  
786 AAGAAATCAACTAGGCC---AGTGTCTAGCCATCTCCCTCTCAGTGGCCCAAACTGA 842  
830 AGACAGGCTCTCAGTTAGATCTTATCAGATTCTTGGACCGACACGACCGCTCTTGC 889  
843 AGACAGGCTCTCAGTTAGATCTTACAGATCTTGGACCGACACGACGCTTGC 902  
890 AATCCAGGAGTGGGAGATACAGCTATGCGAGTTCTTACTGGAGTTCTTGTGGACAG 949  
903 TAATCCAGGTAGTGGCCAGATCCAGCTGTGGCAGTTCTGTCTGAATCTCTGTCCAGACAG 962  
950 CTCCAACCTCCAATGTCATCAGCTGGGAGGACAAATGGGAGTTCAAGATGACAGACCC 1009  
963 CTCCAATCCAACTGATCACTCAGTGGGAGGACCAACGGGAGTTCAAGATGACAGACCC 1022  
1010 TGATGAATGGCTGGCGCTTGGGAGAGAGAGAAAGCAAACTTAATGAATGACTATGACAA 1069  
1023 GGACAGGTGGTTCGGCGCTGGGGGAGAGAGCAAGCCCAACATGAATGACTATGACAA 1082  
1070 ACTCAGCGCTGCACTTCGCTACTACTATGACAAATAATATGACTAAAGTTCAATGTAA 1129  
1083 GCTCAGCGCGCTTCCGCTACTACTACGACAAATAATGATGACGAGGTGACGGGAA 1142  
1130 AGCTATGCTCAAAATTTGATTTCCAGGAATCGCTCAGCGCTCCAGCTCACCCTCC 1189  
1143 GGGCTACGCTTACAGTTTGAATTTCCAGGATTTGCCAGCGCTTGCAGGCCCAACCTCC 1202  
1190 AGAATATCCCATGACAAATACCATCAGACTCCCTTACATGATGTTCTTACCATGACA 1249

1203 TGAGTCTGCTCCCTGTACAAAGTACCCCTCCGACTGCCATACATGGGCTCTATACGCCCA 1262  
1250 CCCCAGAGATGAATTTGTAGTCTCCCAATCCCAATCCCTTGTGCTTCCGTAACCTATCCAG 1309  
1263 CCCCAGAGATGAATTTGTGCTCTCCCAATCCCAATCCCTTGTGCTTCCGTAACCTATCCAG 1322  
1310 CTTTTTGTGCTCCCTTAATCCATCTCATCTGGAATTCACCAACTGGAGCATCTACCCCAATAC 1369  
1323 TTTCTTTGCTTCCCGAACCACATCTGGAATTCACCGACTGGGCGCATCTACCGAACAC 1382  
1370 CAGGTGCCAGTGTCTATATGCTTCCCACTTGGCACTTACTACTA 1417  
1383 TAGGCTCCAGCCCATATGCTCTCTCAGCTGGCACCTACTACTA 1430

RESULT 5  
AY419455 1350 bp DNA linear GSS 17-DEC-2003  
LOCUS Pan troglodytes ERG gene, VIRTUAL TRANSCRIPT, partial sequence,  
DEFINITION genomic survey sequence.  
ACCESSION AY419455  
VERSION AY419455.1 GI:39775412  
KEYWORDS GSS.  
SOURCE Pan troglodytes (chimpanzee)  
ORGANISM Pan troglodytes  
REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.  
AUTHORS Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejarawal, A.,  
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,  
Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Shinsky, J.J.,  
Adams, M.D. and Cargill, M.  
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous  
gene trios  
JOURNAL Science 302 (5652), 1960-1963 (2003)  
PUBMED 14671302  
REFERENCE 2 (bases 1 to 1350)  
AUTHORS Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejarawal, A.,  
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,  
Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Shinsky, J.J.,  
Adams, M.D. and Cargill, M.  
TITLE Direct Submission  
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
Rockville, MD 20850, USA  
COMMENT These sequences were made by sequencing genomic exons and ordering  
them based on alignment.  
FEATURES  
source Location/Qualifiers  
1..1350  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
gene 41..>1350  
/gene="ERG"  
/locus\_tag="HCM6896"  
ORIGIN  
Query Match 59.2%; Score 856; DB 29; Length 1350;  
Best Local Similarity 74.0%; Pred. No. 8e-245;  
Matches 999; Conservative 0; Mismatches 339; Indels 12; Gaps 2;  
QY 81 GAAGCATATCATGTGTGAGTGAAGACCATGCTCTTCTCCAGTGAATATGGCAACATCAAG 200  
DB 1 GAAGCCTTATCATGTGTGAGCGAGGACCATGCTGTTTGTGAGTGTGCTTACGGAACGCCA 60  
QY 141 CACCTTGAACACAGAAATGACAGCTCTCTTCCAGTGAATATGGCAACATCAAG 200  
DB 61 CACCTGGCTAAGACAGAAATGACCGGCTCTCTCCAGCACTATGGACAGACTTCCAG 120  
QY 201 ATGAGCCCGCGGTTCCCAAGCAGAGTGGTTATCACAGCCCCCGGCGCAGATTACCTT 260  
DB 121 ATGAGCCCAACGGTCCCTCAGCAGAGTGGCTGTCTCAACCCCGCAGCCAGGCTCACCATC 180  
QY 261 AAGATGGAGTGTAAACCCAAACCATGTTAATGGTCAAGGAATTCACCTGATGACTGCAGC 320



```
QY 826 CAGAAGACAGGTCCTCAGTTAGATCCTTATCAGATTCCTTGAGCCAGCAGCGGTC 885
Db 129 CAGAAGACAGGTCCTCAGTTAGATCCTTATCAGATTCCTTGAGCCAGCAGCGGTC 198
QY 886 TTGCAATCCAGGAGTGCGGAGATACAGTATGCGAGTTCCTACTGAGCTTCTGTCGG 945
Db 189 TTGCAATCCAGGAGTGCGGAGATACAGTATGCGAGTTCCTACTGAGCTTCTGTCGG 248
QY 946 ACAGCTCCAACTCCAACTGCATCCCTGGGAGGCGCAAAATGGGGAGTTCAAGATGACAG 1005
Db 249 ACAGCTCCAACTCCAACTGCATCCCTGGGAGGCGCAAAATGGGGAGTTCAAGATGACAG 308
QY 1006 ACCCTGATGAAGTGCTCGGCTGGGAGAGAGAGAAAGCAAACTTAACATGACTATG 1065
Db 309 ACCCTGATGAAGTGCTCGGCTGGGAGAGAGAGAAAGCAAACTTAACATGACTATG 368
QY 1066 ACAAACTCAGCGTGCACTTCGCTACTACTATGACAAAATATTATGACTAAAGTTTCATG 1125
Db 369 ACAAACTCAGCGTGCACTTCGCTACTACTATGACAAAATATTATGACTAAAGTTTCATG 428
QY 1126 GTAAAGCTATGCTTACAAATTTGATTTCCAGGAATCGCTCAGGCGCTCAGGCTCAC 1185
Db 429 GTAAAGCTATGCTTACAAATTTGATTTCCAGGAATCGCTCAGGCGCTCAGGCTCAC 488
QY 1186 CTCAGATCATCCATGTACAAATACCCATCAGACCTCCCTTACATGATGTTCTTACCATG 1245
Db 489 CTCAGATCATCCATGTACAAATACCCATCAGACCTCCCTTACATGATGTTCTTACCATG 548
QY 1246 CACACCCCGAGAGATGAATTTGAGTCCCATCCCTGCTTTCCCGGTAACTTCAT 1305
Db 549 CACACCCCGAGAGATGAATTTGAGTCCCATCCCTGCTTTCCCGGTAACTTCAT 608
QY 1306 CCAGCTTTTGTGCTCCCTTAATCCATCTCGAATTCACCAACT-GGAGGCATCTACCCC 1364
Db 609 CCAGCTTTTGTGCTCCCTTAATCCATCTCGAATTCACCAACTGGAGGCATCTACCCC 669
QY 1365 AATACGAGCTGCCAGCTGCATATGCTTCCCATC-TTGGCAGCTACTACTAAGTG 1421
Db 669 AATACGAGCTGCCAGCTGCATATGCTTCCCATC-TTGGCAGCTACTACTAAGTG 726

RESULT 7
LOCUS AK042111
DEFINITION Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched
library, clone:A630059111 product:Friend leukemia integration 1,
full insert sequence.
ACCESSION AK042111
VERSION AK042111.1 GI:26334936
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
```

```
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,S., Watahiki,M.,
Yoneda,Y., Ishikawa,T., Tanaka,K., Tanaka,T., Matsura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multichannel sequencer
Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium
TITLE Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team
TITLE Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 6
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,M.,
Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shingawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaki-Akahira,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
Trust/MRC Building Addenbrookes Hospital Cambridge) whose
assistance we gratefully acknowledge.
Please visit our web site for further details.
URL:http://genome.gsc.riken.go.jp/
URL:http://fantom.gsc.riken.go.jp/
FEATURES
Location/Qualifiers
1..1706
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:A630059111"
/db_xref="MGI:2407136"
/db_xref="taxon:10090"
/clone="A630059111"
/tissue_type="thymus"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="3 days neonate"
158..1516
/notes="unannotated protein product; Friend leukemia
integration 1 (MGI:95554, GB|NM_008026, evidence:
BLASTN, 99%, match=1671)
putative"
/codon_start=1
/protein_id="BAC31169.1"
/db_xref="GI:26334937"
CDS
```

/translation="MDGTIKKALSVVSDSDSLFDSAYGAAHLPKADMTASGSPDYQG  
PKINLPPOEMINOPVNVVREYDHMGSRSPVDCSVSKCNKLVGGSEANPMVY  
NSYMDKXGPPBNMTNERRVUPADPTLQTOEHVROKLEWAKIKEYGLMEIDTSFO  
NMCKELCKXNKEDFLRATSAYNTEVLLSHLSYLRSSLLAYNTTSHITQSSRLNKE  
DPSYDSVRGAWNNMNSGLNKPLIGSSQTMGNTKEQRPDPYQI.LGPTSLRLANP  
GSGOIQWLQFLLELLSDSANASCTIWEITGTFBFTMDPDEVARWGERKSPNMYDK  
LSRLALFYVFNKIMTKYHGRYAKFDFHGIQAOLQPHPTTSMYKPSDISYMP5YH  
AHQKRVFVPSHPSSMBVTSSSPFGAASQYWTSPTAGIYPNPSVPRHENTHVP5HLS  
YY"

ORIGIN

Query Match	39.4%; Score 570.2; DB 11; Length 1706;
Best Local Similarity	66.8%; Pred. No. 5e-159;
Matches 917; Conservative 0; Mismatches 423; Indels 33; Gaps 6;	
QY	63 ATGCAAGCAGCATTATTAAGGAGCATTATCAGTGTGTGAGTGAAGACCAAGTCCCTGTTTGAG 122
DB	158 ATGACGGGACTATTAGGAGCTCTCTGTGTGTGTGAGTGCAGATCACTCCCTTTTGAT 217
QY	123 TGTGCTTACGGATCG----CCCCCTTGCAGAAAGACAGAAATGACAGCCTCTCTTCCAGT 179
DB	218 TCAGATACGGAGCGGAGCCCATCTCCCAAGGCAGATATGACTGCTTCGGGAGTCCCT 277
QY	180 GAATATGGGAAAACATCAAGATCAGCCCGCGCTTCCCAGCAGGACTGTTATCAGAG 239
DB	278 GACTACGGGAGCCCAAAATCAACCCCTGCGACCGCAGCAGGAGTGGATCAACCAG 337
QY	240 CCCCAGCCAGAGTTACCATTAAGATGGAGTGTAAACCCAAACAGGTTAATGGGTCAAGG 299
DB	338 CC-----AGTGAGAGTCAATGTCAAGCGGAGTATGACACATGAATGGATCCAGG 388
QY	300 AATTCACTGTATGACTGAGCGTGGCAAGGAGGAGGAATGTTAGCAGTTCAGACAAT 359
DB	389 GAGTCTCCGGTGGAGTCAAGTGTGAGCAATGTAAACAGCTGTGGCGGAGGAGGCC 448
QY	360 GTTGGGATGAATGGAAGCTACATGAAGAGAGCA---TATTCCGCTCCAAATATG 416
DB	449 AACCCATGACTATATAGTCTATGATGAGAGAGACGGCCCCCTCTCTCCCAACATG 508
QY	417 ACAACCAATGAAGAGAGTTATTGTGCGCAGCAGATCTTACGTTATGAGGACAGACCAT 476
DB	509 ACCACCAAGCAAGGAGAGTCAATTGTGCTGTCAGACCCCACTGTGGACACAGGAGCAC 568
QY	477 GTACGGAGTGGCTGGAGTGGCAGTGAAGAGTATGCTTTCAGACGCTGGACATCTTG 536
DB	569 GTTCAAGTGGCTGGAGTGGCTATTAAGAAATACGATGAGTGGAGATGACACTTCC 628
QY	537 TTGTTCCAGAACATGTATGGGAAAGAGTTGTGTAATAATGACCAAGATGACTTCCAGAGA 596
DB	629 TTCCTCCAGAACATGTATGGTGAAGCAATGTGTAATAATGAACAGGAGGAGTCTCTCCGA 688
QY	597 CTCAGCGCAGCTATACGACAGATATCTCTGTCACACCTACACTACCTCAGAGAGAGA 656
DB	689 GCCACCTCCGCTCAACACAGAGAGTGTGTGTGCGACCTCAGTTACCTCAGGGAAG- 747
QY	657 GGAGCCACTTTTATTTTCCAAATACATCAGTTTACCCAGAGCAACGCAAGAAATACA 716
DB	748 -----TTCACCTGCTGGCCTATAACACAACTCCCATACAGACCAAGTCTCTCAGCACTGAAT 802
QY	717 ACAAGCCAGATTTACCTTATGACCAAGCAGGAGATCAGCGTGGAGGAGTCAACA--GC 773
DB	803 GTCAGGAGAGCCCTTCTTATGACTCTCTCAGGAGAGGAGATGGAACAATATATGAAC 862
QY	774 CATCCCACTCAGTCAAAAGCTACCAACCATCATCTTCAACAGTGCACCAACAGAGAGAC 833
DB	863 TCTGGCCTCAACAAAGCTCTCTCTCTGAGGATCAGACACCATGGGCAAGAACACTGAG 922
QY	834 CAGCCTCTCAGTTAGATCTTATCAGATTTCTTGGACCGACAGCGCGCTTGGCAAT 893
DB	923 CAGCGGCCCAAGCCAGATCTTATCAGATCTCTGGGGCCCAACAGCAGCGCGCTAGCAAC 982
QY	894 CCAGGAGTGGGCAGATCAGTATGGCAGTCTCTACTTGGAGCTTCTGTGCGACAGCTCC 953

DB	983 CCTGGAGTGGGAGATCCAGCTGTGGCAGTTTCTCTCGGAACACTACTGTCCGACAGCC 1042
QY	954 AACTCCAACTGCATCACCTGGGAGGGCAAAATGGGAGTTTCAAGATGACAGACCTGAT 1013
DB	1043 AACGCCAGCTGTATCACCTGGGAGGGACCAACGGGAGTTTCAAAATGACGAGCCCTGAT 1102
QY	1014 GAAGTGGCTGCGCGTGTGGGAGAGAGGAAAAGCAAACTTAACATCAACTATGACAAATC 1073
DB	1103 GAGTGGCCAGCGCTGTGGGAGAGCGGAAGAGCAAGCCCAACATGAATATGACAGCTG 1162
QY	1074 AGCGTGACCTTCGCTACTACTATGACAAAATATATGACTTAAAGTTTCATGTTAAAC 1133
DB	1163 AGCCGGGCGCTCCGATACTACTATGACAAAACATATTAAGCAAAAGTGTGCGAAGG 1222
QY	1134 TATGCTTACAAATTTGATTTCCACGGAATCGTTCAGGGCCCTCCAGCCTCACCTCCAGAA 1193
DB	1223 TATGCTTACAAATTTGATTTCCACGGAATCGTTCAGGGCCCTCCAGCCTCACCAAGAG 1282
QY	1194 TCATCCATGTACAAATACCATCCATCAGACCTCCCTACATGAGTTCTCCATCAGCAGCC 1253
DB	1283 ACATCCATGTACAGTATCCCTCTGATATCTCTACATGCGCTTCTTACATGCCATCAA 1342
QY	1254 CAGAAGATCAATTTGTAGTCCCATCCCTGCTTTTGGCCGTAACTCATCTCCAGCTTT 1313
DB	1343 CAGAAGTGAATTTGTCCGCTCTCAGCCATCTCCATGCTGTCACTCTCCAGCTTC 1402
QY	1314 TTTGCTGCCCTTAATCCATCTGGAATTCACCACTGAGGAGCTTCCATCCATACCAGG 1373
DB	1403 TTTGGAGCAGCATCAATCTGAGCTCCCTCCATCTGCTGGGATCTATCCAAACCCAGT 1462
QY	1374 CTGCCAGCTGCTCATAT-----TGCTTCCCATCTTTGGCACCTACTACTA 1417
DB	1463 GTCCCCCGCATCTTAACACCACCGTGCCTTACACTTAGGAGCTACTACTA 1515

RESULT 8  
AY418895 1347 bp DNA linear GSS 17-DEC-2003  
LOCUS Mus musculus FL11 gene, VIRTUAL TRANSCRIPT, partial sequence,  
DEFINITION genomic survey sequence.  
ACCESSION AY418895  
VERSION AY418895.1 GI:39774855  
KEYWORDS GSS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,  
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,  
Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,  
Adams, M.D. and Cargill, M.  
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous  
gene trios  
JOURNAL Science 302 (5652), 1960-1963 (2003)  
PUBMED 14671302  
REFERENCE 2 (bases 1 to 1347)  
AUTHORS Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,  
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,  
Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,  
Adams, M.D. and Cargill, M.  
TITLE Direct Submission  
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
Rockville, MD 20850, USA  
COMMENT These sequences were made by sequencing genomic exons and ordering  
them based on alignment.  
FEATURES  
Location/Qualifiers  
source 1..1347  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
c1..>1347  
/gene="FL11"



```

ORIGIN
Query Match      38.7%; Score 560.6; DB 29; Length 1347;
Best Local Similarity 66.7%; Pred. NO. 3.4e-156;
Matches 905; Conservative 0; Mismatches 419; Indels 33; Gaps 6;

QY 79 AGGAAGCATTTATCAGTGTGAGTGAAGACCAAGTCCCTGTTGTTGAGTGTGCTTACGGATCGC 138
DB 5 AGGAGGCTCTGTCTGTGTGTGAGTGAAGATCAGTCCCTTTTGTATCAGCATACGGAGCGG 64
QY 139 ---CCACCTTGCAGAAAGACAGAAATGACAGCCTCTCTTCCAGTGAATATGAGGCAACAT 195
DB 65 CAGGCCATCTCCCAAGGAGATATGACTGCTCGGGAGTCTCTGACTAGCGGAGGCC 124
QY 156 CAAGATCAGCCGCGCTTCCACAGAGGACTGTTATACAGCCCGCGCCAGAGCTTA 255
DB 125 ACAAAATCAACCCCTGCGACGAGAGAGTGAATCAACAGCC-----AGTGA 175
QY 256 CAATTAAGATGAGTGTAAACCAACAGGTTAATGGGTCAAGAAATCACTGATGACT 315
DB 176 GAGTCAATGTCAAGCGGAGTATGACCACATGAATGATCCAGGAGTCTCCGGTGGACT 235
QY 316 CGAGCGTGGCAAGAGGAGAAATGTTAGCAGTTTCAAGATTCAGACATGTTGGGATGAACATG 375
DB 236 GCGTGTGAGCAATGTAAACAGCTGTGTGGCGGAGGCGAGCCAAACCCATGACTATA 295
QY 376 GAAGCTACATGAGAGAGAACCA---TAATCCGCTCCAAATATGACAAACCAATGAACGAA 432
DB 296 ATAGCTACATGATGAGAAGAACGGCCCTCTCTCCCAACATGACCAACCAACGGA 355
QY 433 GAGTTATTGTGCCACAGATCCTAGTTATGGAGCAGACATGATGAGGAGTGGGTGG 492
DB 356 GAGTCATTGTGCTCGACACCCACACTGTGTGACACAGAGGACGTTGCAAGTGGGTGG 415
QY 493 AGTGGGAGTGAAGGATGTGTCTTCAGAGCTGGACATCTTGTGTTCCAGAACATTTG 552
DB 416 AGTGGGCTATAAGAAATACGAAATGATGGAGATTGACACTCTCTTCCAGAACATGG 475
QY 553 ATGGGAAGAGTGTGTAAATGACCAAGATGACTTCCAGAGACTCAGCCGAGCTATA 612
DB 476 ATGGCAAGGAATGTGTAAATGAAACAGGAGGACTTCTCCGAGCCACCTTCCGCTTACA 535
QY 613 AGCAGATATCTCTCTGTACACCTTACACTACCTCAGAGAGAGGAGGACCACTTTTATT 672
DB 536 ACACAGAAGTGTGTGTGACCTTACCTTACCTCAGGGAAG-----TTCACGTGTGG 589
QY 673 TTCCAAATACATCAGTTTACCCAGAACGACGCAAGAAATACAAAGGCCAGCATTTAC 732
DB 590 CCTATAACACAACTCCCATACAGACCAAGTCTCTCAGACTGAATGTCAAGAGAGACCTT 649
QY 733 CTTATGACAGCAGGAGAGATCAGCGTGGACGAGTCAAGCCATCCCACTCAGTCAAAAG 792
DB 650 CTTATGACTCTGTCTCAGGAGAGAGATGGAACAATAATGACTCTGGCCCTCAACAAA 709
QY 793 CTACCCAAACCATCATCTTCA---ACAGTGGCCCAAAACAGAGACAGCGTCTCAGTTAG 849
DB 710 GTCCCTCTCTTGGAGGATCACAGACCATGTGGCAAGAACACTGAGCAGCGGCCCGAGCAG 769
QY 850 ATCCTTATCAGTTCTTGGACGACAGCAGCGCTTTCGAAATCCAGGAGTGGGAGA 909
DB 770 ATCCCTATCAGATCTCTGGGCGCAACAGAGCGGCTTAGCAAACTCTGGAGTGGGAGA 829
QY 910 TACAGCTATGGCAGTCTCTTCTGAGCTTCTGTGGACAGCTTCCAACTTCCAACTGCA 969
DB 830 TCCAGCTGTGGCAGTTCTCTCTGGAACACTGTCTCGACAGCGCCAAACCCAGCTGTATCA 889
QY 970 CTGGGAGGGCAAAATGGGAGTTCAAGATCAGACGCTGATGAAGTGGCTCGGGTT 1029
DB 890 CTGGGAGGGGCAACAGCGGAGTTCAAATGACGGACCTTGTGAGTGGCCAGGCGCT 949
QY 1030 GGGGAGAGGAGAAAGCAAACTTAACATCAACTATGACAACTCAGCGCTGCATCTCGCT

```

```

DB 950 GGGGAGAGCGGAAGAGCAAGCCCAACATGAATTATGACAAGCTGAGCGGCGCTCCGAT 1009
QY 1090 ACTACTATGACAAATATATTGACTTAAGTTCATGGTAAAGCTATGCTTCAACAATTG 1149
DB 1010 ACTACTATGACAAATATATTGACTTAAGTTCATGGTAAAGCTATGCTTCAACAATTG 1069
QY 1150 ATTTCCACGGAATCGCTCAGGCCCTCCAGCCTCCACCTCCAGAAATCATCCATGTACAAT 1209
DB 1070 ACTTCCATGGCATTGCCAGGCCCTTGACCCACATCCCAACAGAGACATCCATGTACAAGT 1129
QY 1210 ACCATCAGACCTCCCTTACATGAGTTCTTCTACCATGCAACCCCGAGAGATGAACTTTG 1269
DB 1130 ATCCCTCTGATATCTCTTACATGCTTCTACCATGCCCCATCAACAGAGAGTGAATTTG 1189
QY 1270 TAGCTTCCCATCCCTCTGCTTTGCCCGTAACTCATCCAGCTTTTTCGTGCCCTTAATC 1329
DB 1190 TCCCGTCTCAACCATCTCTCATGCTGTCCACCTCTCCAGCTTCTTTGGAGCAGCATCAC 1249
QY 1330 CATCTGGGAATTCACCACTGGAGCCTTCTCCCAATACCACTGCGCTGCTCATATA 1389
DB 1250 AATCTGGACCTTCCCGCACTGTCTGGATCTATCCAAACCCAGTGTCCCGCCCATCTTA 1309
QY 1390 -----TGCTTCCCATCTTGGCACCCTACTACTA 1417
DB 1310 ACACCCACGTCGCTTACACTTAGCAGCTACTACTA 1346

RESULT 9
AK036655 3118 bp mRNA linear HTC 19-SEP-2003
LOCUS Mus musculus adult male bone cDNA, RIKEN full-length enriched
DEFINITION library, clone:9830146N17 product:Friend leukemia integration 1,
full insert sequence.
ACCESSION AK036655.1 GI:26331575
VERSION HTC; CAP trapper.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636
PUBMED
REFERENCE
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159
PUBMED
REFERENCE
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
PUBMED
REFERENCE
4 The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

```



**AUTHORS** The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
**TITLE** Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
**JOURNAL** Nature 420, 563-573 (2002)  
**REFERENCE** 6 (bases 1 to 3118)  
**AUTHORS** Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayasaka, T., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katch, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuura, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Oheato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toyu, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)  
**COMMENT** cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/.  
**FEATURES** Location/Qualifiers  
source  
1..3118  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="CS7BL/6J"  
/db\_xref="FANTOM\_DB:9830146N17"  
/db\_xref="MG1:2401092"  
/db\_xref="taxon:10090"  
/clone="9830146N17"  
/sex="male"  
/tissue\_type="bone"  
/clone\_libs="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
192..1547  
/feature="unnamed protein product; Friend leukemia integration 1 (MGD|MG1:95554, GB|NM\_008026, evidence: BLASTN, 99%, match=1671)  
putative"  
/codon\_start=1  
/protein\_id="BAC29518.1"  
/db\_xref="GI:26331576"  
/translation="MDCITKEALSVVSDQSLFDSAYGAALPKADMTASGSDYGG  
PHKINLPQCEWNTQVVRVNEVDHMGSPSPVCSVKCNKLVGGNARNPMYN  
SIMDKNGPFPNNMTNRRVVPADPTLWQHVQRLNLEAKYGLMEDTSTFFQK  
MDGKELCKNKKEDFIRATSYNTLVLLSHLSLURESSLAYINTSTTQSSRLNVRK  
PSYSDVRGANNNSGNKSLPLGSGQTMGNQKQRFQPDPPYQILGPTSSRLANPK  
SQIQWQFLLELSDSANGCTTWECTNGEFTNTDPEVARWGERKS KPNMYNDKI  
SRALRYVYDKNMTKVKGYAYKDFPHGLAQALQPHPTETSMYKYPSPDISYSPSYHA  
HQQKNVFPVPSHPSSMPVTSSSFFSGAASQYVTSPTAGIYDNPSPVRPHTHPVSHLGS  
Y"  
**ORIGIN**  
Query Match 38.6%; Score 559; DB 11; Length 3118;  
Best Local Similarity 66.8%; Pred. No. 1.5e-155;  
Matches 917; Conservative 0; Mismatches 420; Indels 36; Gaps 7  
  
QY 63 ATGCGAAGCACTATTAAAGGAGCATTATCAGTGGTGGTGAACACGAGCTCTTTGAG 122  
DB 192 ATGCGAGGCACTATTAAAGAGCTCTGCTGTGTGGTGGTGAACATCAGTCTTTTIGAT 251

```
QY 1194 TCATCAATGTAATAATACCATCAGACCTCCCTACATAGTTCCTACCATGCACACCCC 1253
Db 1314 ACATCCATGTACAGTATCCCTCTGATATCTCTACATGCTTCTACCATGCCATCAA 1373
QY 1254 CAGAAGATGAATTTGTAGTCCCATCCCTCCCTGTTGCGCGTAACTCATCCAGCTTT 1313
Db 1374 CAGAAGGTGAATTTGTCCCGTCTCACCCATCTCCCATGCTGTCACCTCTCCAGCTTC 1433
QY 1314 TTGTGTCGCCCTAATCCATATCGGAATTCACCAACTGGAGGCATCTACCCCAATACCAGG 1373
Db 1434 TTGGAGCAGCATCAATATATGACCTCCCTCCCTGCTGGATCTATCCAAACCCCATG 1493
QY 1374 CTGCCAGTGTCTATA-----TGCTTCCCATCTTGGCAGCTACTACTA 1417
Db 1494 GTCCCCCGCCATCTTAACACCCACGTGCTTTCACACTTAGGCAGCTACTACTA 1546

RESULT 10
LOCUS AY418893 1347 bp DNA linear GSS 17-DEC-2003
DEFINITION Homo sapiens FLI1 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY418893
VERSION AY418893.1 GI:39774853
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1347)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1347)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT These sequences were made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source Location/Qualifiers
1..1347
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>1347
/gene="FLI1"
/locus_tag="HCM6709"

gene

ORIGIN
Query Match 38.2%; Score 552.6; DB 29; Length 1347;
Best Local Similarity 66.3%; Pred. No. 8.5e-154;
Matches 900; Conservative 0; Mismatches 424; Indels 33; Gaps 6;

QY 79 AGGAAGCATTTATCAGTGGTGAAGACAGTCCCTTTGTTGAGTGTGCTACGGATCGC 138
Db 5 AGGAGGCTCTGTGCGGTGGTGAAGACAGTCCCTTTGACTCAGCGTACGGAGCGG 64
QY 139 ---CCACCTTGAAGACAGAAATGACAGCTCCCTCTCCAGTGAATATGGGCAACAT 195
Db 65 CAGCCCATCTCCCAAGGCGGACATGACTGCTTCGGGGAGTCTGACTACGGGCGGCC 124
QY 196 CAAAGATGAGCCCGCGTTCCTCCAGCAGGACTGGTTATCAGACCCCGGCCAGAGTTA 255
Db 125 ACAAGATCAACCCCTCCCAACACACAGCAGAGTGGATCAATCAGCC-----AGTGA 175
```

```
QY 256 CCATTAAGATGGAGTGTAAACCCAAACCCAGGTTAATGGGTCAAGGAATTCACTGATGACT 315
Db 176 GGGTCAACGCTCAAGCGGAGTATGACCACATGAATGGATCCAGGGAGTCTCCGGTGGACT 235
QY 316 GCAGCGTGCAAAAGGAGGAAATGGTTAGCAGTTTCAGACAATGTGGGATGAACATATG 375
Db 236 GCAGCGTTAGCAAAATGCGAGCAAGCTGGTGGCGGAGGCGAGTCCAAACCCCATGACTACA 295
QY 376 GAAGCTTACATGAAGAGAGAGCAT---ATTCCGCCTCCAAATATGACAACCAATGAACGAA 432
Db 296 ACGCTATATGACGAGAGAAATGCCCCCTCTCCCAACATGACCAACCAACGAGAGGA 355
QY 433 GAGTTATTGTGCGCAGCAGATCTCTAGCTTATGAGCAGACAGCACCATGTAGCGCAGTGGCTGG 492
Db 356 GAGTCATCTCTCCCGCAGACCCCACTGTGTGACACAGGAGCATGTGAGGCAATGGCTGG 415
QY 493 AGTGGCAGTGAAGGAGTATGGTCTTCCAGACGTGGACATCTTGTGTTTCCAGAACATTTG 552
Db 416 AGTGGGCCATAAGGAGTACAGCTTGATGGAGATCGACACATCTCTTTTCCAGAACATGG 475
QY 553 ATGGGAAAGAGTTGTGTAATAATGACCAAGATGACTTCCAGAGACTCAGCGCGAGCTATA 612
Db 476 ATGGCAAGGAACTGTGTAAATGAACAAGGAGGACTTCTCCGCGCCACCAACCTCTACA 535
QY 613 ACGCAGATATCTCTCTGTCTACACCTACACTCTCAGAGAGAGAGAGCCACTTTTATTT 672
Db 536 ACACGGAAGTGTGTTGTTCACCTCAGTTACTCAGGGAAG-----TTCACGTCTGG 589
QY 673 TTCAAATATCATCAGTTTATCCAGAAAGCAACGCAAGAATAACAAACAGGCCAGATTAC 732
Db 590 CCTATAATACAACTCCCAACACCGCAATCTCTCAGATTGATGTGTCAAAGAAGACCCCTT 649
QY 733 CTTATCAGCAACGCGAGGATCAGCGTGCAGCAGTCCAGCCTCCCTCAGTCAAAA- 791
Db 650 CTTATGACTCAGTCAGAGAGGAGCTTGGGCAATACATGATTTCTGSCCTCAACAAA 709
QY 792 --GCTACCCCAACCATCATCTTCAACAGTGCCTCCAAACAGAAAGCAGCGTCTCAGTTAG 849
Db 710 GTCTCTCCCTTGGAGGGGCACAAACGATCAGTAAGAATAACAGAGCAACGGCCCCAGCAG 769
QY 850 ATCTTTATCAGATCTTTCAGCCAGCAGCAGCGCTCTTCAAAATCCAGGGAGTGGCGAGA 909
Db 770 ATCCGTATCAGTCTCTGGGCCACAGCAGTCCCTAGCCCTAGCCACCTTGGAGCGGCGAGA 829
QY 910 TACAGCTATGGCAGTTCCTACTGAGCTTCTGTGGAAGCTCCAACTCCAACTGATCA 969
Db 830 TCCAGCTGTGGCAATTCCTCTGAGCTGCTCTCCGACAGCGCAACGCCAGCTGTATCA 889
QY 970 CTTGGGAGGCGCAAAATGGGAGTTCAAGATGACAGACCTCGATGAAGTGGCTCGCGCTT 1029
Db 890 CTTGGGAGGCGCAACCGGGAGTTCAAAATGACGGACCCCGATGAGTGGCCAGCGCT 949
QY 1030 GGGGAGAGAGAAAAGCAAACTTAACATGAACCTATGACAACTCAGCCCTGACATCTCGCT 1089
Db 950 GGGGAGGCGGAAAAGCAAGCCCAACATGAATTACGACAGCTGAGCGGGCCCTCGCTT 1009
QY 1090 ACTACTATGACAAAATATTATGACTTAAGTTCAATGGTAAACCTATGCTACAAATTTG 1149
Db 1010 ATTACTATGATAAAACATTTATGACCAAAAGTGCACGGCAAAAGATATGCTTTACAAATTTG 1069
QY 1150 ATTTCCACGGAATCGGTTCAGGCGCTCCAGCTCCAGCTCCAGAAATCATCTCATCAAAAT 1209
Db 1070 ACTTCCACGCAATGCGCGAGCTCTCGAGCCACATCGAGCCGAGTCTGCTCATCAAGT 1129
QY 1210 ACCCATCAGACCTCCCTCATCATGAGTTCTTACATGCAACCCCGCAGAAATGAACTTTG 1269
Db 1130 ACCCTTCTGACATCTCTCTACATGCTTCTCTACCATGCTCCCAACGAGGAGGTGAACCTTTG 1189
QY 1270 TAGCTCCCATCCCTGCTTTGCGGTAACTCTCATCCAGCTTTTGTGCTGCCCTTAATC 1329
Db 1190 TCCCTCCCATCCCATCTCCATGCTGCTCACTTCTCTCCAGCTTCTTTGGAGCGCATCAC 1249
```

```

QY 1330 CATACTGGAATTCACCAACTGGAGGCAATCTACCCCAATACA-----GGCTGCCAG 1380
Db 1250 AATACTGGACCTCCCCACGCGGGGAATCTACCCCAACCCCAAGCTCCCGCCGATCCTA 1309
QY 1381 CTGCTCATATGCTTCCCATCTTGGCACCTACTACTA 1417
Db 1310 ACACCCAGTGCCTTCACACTTAGCAGCTACTACTA 1346

RESULT 11
BU478556/c 606 bp mRNA linear EST 30-NOV-2002
LOCUS 603471895F1 CSEQ8BN22 Gallus gallus CDNA clone CHEST349m24 5', mRNA
DEFINITION sequence.
ACCESSION BU478556
VERSION BU478556.1 GI:25972133
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
1 (bases 1 to 606) Phasianinae; Gallus.
Fong, W.T., Tickle, C., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
BOARDMAN, P.E., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
22335534
MEDLINE
PUBMED 12445392
COMMENT Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

FEATURES
source
1. 606
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Layer and broiler"
/db_xref="taxon:9031"
/clone="CHEST349m24"
/sex="Male and female"
/tissue_type="Chondrocytes isolated from growth plate
cartilage"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="CSEQ8BN22"
/note="Vector: pBluescript II KS(+); Site_1: EcoRI;
Site_2: NotI; This normalized library was constructed from
1 million independent clones. cDNA synthesis was initiated
using an oligo(dT) primer, using methylated C in the first
strand synthesis reaction. Following this first strand
reaction, double-stranded cDNA was blunted, ligated to
NotI adapters, digested with EcoRI, size-selected, and
cloned into the NotI and EcoRI compatible sites of a
custom modified MCS of the pBluescript (KS+) vector. The
library was normalized in 2 rounds using conditions
adapted from Soares et al., PNAS (1994) 91: 9228-9232 and
Bonaldo et al., Genome Research 6 (1996): 791, except that
a significantly longer reannealing hybridization was
used."

ORIGIN
Query Match 36.4%; Score 526.2; DB 13; Length 606;
Best Local Similarity 97.5%; Pred. No. 4.8e-146;
Matches 545; Conservative 0; Mismatches 13; Indels 1; Gaps 1;

QY 707 AAGGATTACCAAGGCGCAGATTTCCTATGACCAAGCGAGGAGATCACCGTGGACGAG 766
Db 590 AAGGATTACCAAGGCGCAGATTTCCTATGACCAAGCGAGGAGATCACCGTGGACGAG 531

```

```

QY 767 TCACAGCCATCCCACTCACTAGTCAAA-AGCTACCAACCACTCATCTTTCAACAGTGCACAAA 825
Db 530 TCACAGCCATCCCACTCACTAGTCAAAAGAGTACCCCAACCACTCATTTTCAACAGTGCACAAA 471
QY 826 CAGACAGCAGCGTCTCAGTGTAGTATCTTATCAGATTCTTGGACCGACCGACCGC 885
Db 470 CAGAAGACCGAGCGTCTCAGTGTAGTATCTTATCAGATTCTTGGACCGACCGACCGC 411
QY 886 TTGCAAAATCCAGGAGTGGGAGATACAGCTATGCGAGTTCCTACTGGAGCTTCTGTGG 945
Db 410 TTGCAAAATCCAGGAGTGGGAGATACAGCTATGCGAGTTCCTACTGGAGCTTCTGTGG 351
QY 946 ACAGCTCCAACTCCAACTGCATCACTCGGAGGGGACAAATGGGAGTTCAGATGACAG 1005
Db 350 ACAGCTCCAACTCCAACTGCATCACTCGGAGGGGACAAATGGGAGTTCAGATGACAG 291
QY 1006 ACCCTGATGAGTGGCTCGGCTGGGAGAGAGAGAAAGCAAACTAACATGAACATG 1065
Db 290 ACCCTGATGAGTGGCTCGGCTGGGAGAGAGAGAAAGCAAACTAACATGAACATG 231
QY 1066 ACAACTCAGCGCTGCATCTCGCTACTACTATGACAAAAATATTATGATAAAGTTCATG 1125
Db 230 ACAAACTCAGCGCTGCATCTCGCTACTACTATGACAAAAATATTATGATAAAGTTCATG 171
QY 1126 GTAAACGCTATCGCTACAAATTTGATTTCCAGGGAATCGCTCAGGCCCTCCAGCGCTCACC 1185
Db 170 GTAAACGCTATCGCTACAAATTTGATTTCCAGGGAATCGCTCAGGCCCTCCAGCGCTCACC 111
QY 1186 CTCGAGAATCATCATGTACAAATACCATCAGACCTCCCTTACATGATGTTCTTACCATG 1245
Db 110 CTCGAGAATCATCATGTACAAATACCATCAGACCTCCCTTACATGATGTTCTTACCATG 51
QY 1246 CACACCCCGCAGAGATGAA 1264
Db 50 CACACCCCGCAGAGATGAA 32

```

```

RESULT 12
AY418894 1256 bp DNA linear GSS 17-DEC-2003
LOCUS Pan troglodytes FLII gene, VIRUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY418894
VERSION AY418894.1 GI:39774854
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
1 (bases 1 to 1256)
Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarawal, A.,
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
Perriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
Adams, M.D. and Cargill, M.
Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
2 (bases 1 to 1256)
Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarawal, A.,
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
Perriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
Adams, M.D. and Cargill, M.
Direct Submission
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
These sequences were made by sequencing genomic exons and ordering
them based on alignment.
Location/Qualifiers
1. 1256
/organism="Pan troglodytes"
/mol_type="genomic DNA"

```

QY	1213	CATCAGACCTCCOCTACATGAGTTCTTACCATGCACACCCCGAGAGATGAATTTGTAG	1272
Db	1042	CTTCTGACATCTCTCATATGCTTCTTACCATGCCACACGAGAGGTGAATTTGTCC	1101
QY	1273	CTCCCATCCCTGCTTTGTCGCGTAACTTCATCCAGCTTTTGTGTCGCCCTAAATCCAT	1332
Db	1102	CTCCCATCATCTCTCATGCGCTGTCACTCTCTCCAGCTCTTTTGGAGCCGCATCAAT	1161
QY	1333	ACTGGAATTCACAACTGGAGGATCTACCCCAATACCA-----GCTGCCAGCTG	1383
Db	1162	ACTGGACCTCCCCANNNGGGAATCTACCCCAACCCACGTCGCCGCATCCTAACCA	1221
QY	1384	CTCATATGCTTCCCATCTTGGCACTACTACTA	1417
Db	1222	CCCACGTGCTTCCACACTTAGGCAGCTACTACTA	1255
RESULT 13			
LOCUS	BQ233264	880 bp	mrna linear EST 02-MAY-2002
DEFINITION	AGENCOURT_7565936 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6041526		
ACCESSION	5', mRNA sequence.		
VERSION	BQ233264		
KEYWORDS	BQ233264.1 GI:20414664		
SOURCE	EST.		
ORGANISM	Homo sapiens (human)		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 880)		
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: sgabs-@email.nih.gov Tissue procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arranged by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LHAM13279 row: h column: 07 High quality sequence stop: 608. Location/Qualifiers 1. 880 /organism="Homo sapiens" /mol_type="mrna" /db_xref="taxon:9606" /clone="IMAGE:6041526" /tissue_type="embryonal carcinoma, cell line" /lab_host="DH10B (phage-resistant)" /clone_lib="NIH MGC 92" /notes="Organ: testis; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."		
FEATURES			
source	1. 880		
ORIGIN			
Query Match	33.5%;	Score 485.4;	DB 13; Length 880;
Best Local Similarity	84.4%;	Prod. No. 1e-133;	
Matches 546;	Conservative	0;	Mismatches 101; Indels 0; Gaps 0;
QY	8	CGGACGAATAATTTATTAGCAATTTATTAGCATCAATAATCTTGATCAATATGGC	67
Db	19	GAGAAGCAATAATTTATTAAATTTATTAAAGATCAATAAACTTGATTCATTTGGC	78
QY	68	AAGCACTATTAAAGAGCAATTTACGTGGTGAAGACCAAGTCCTTTGTTGAGTGTGC	127
Db	79	CAGACCTATTAAAGAACCTTATCAGTTGTGAGTGAGACCAAGTCGTTGTTGAGTGTGC	138

QY 128 CTACGGATCGCCCACTTGGAAAGACAGAAATGACAGCTCTCTCTCCAGTGAATATGG 187  
DB 133 CTACGGAACGCGACACCTTGGCTTAAGACAGAGATACCGCGTCTCTCCAGCGACTATGG 198  
QY 188 GCAACATCAAGATGAGCGCGGTTCCCGAGCAGGACTGGTTATCACAGCCCGCGC 247  
DB 199 ACAGACTTCCAGATGAGCCCAAGGTCCTTCAGCAGGATGGGTCTCTCAACCCCGAGC 258  
QY 248 CAGAGTTACCTTAAGATGGAGTGAACCCAAACAGAGTTAATGGTCAAGGAATTCACC 307  
DB 259 CAGGCTACCATCAAAATGGAATGTAACCTTAGCCAGGTGAATGGCTCAAGGAATCTCC 318  
QY 308 TGATGACTGACGCGTGGCAAGAGGGAATGTTAGCAGGTTTCAGACAATGTTGGAT 367  
DB 319 TGATGAATGCAAGTGGCCAAAGCGGGAAGATGGTGGGAGCCAGACACCGTTGGAT 378  
QY 368 GAACATATGGAAGCTACATGGAAGAGAGCATATTCGCGCTCCAAATATGACAACTGA 427  
DB 379 GAACTACGGCAGTACATGGAAGAGAGCAGATGCGACCCCAACATGACACGAAAGCA 438  
QY 428 ACGAAGATTTATGCGCAGCAGATCTTACCTTATGAGACAGACACCATGTACGGCAGTG 487  
DB 439 GCGCAGATTTATGCGCAGCAGATCTTACCTATGGAATGGAATGGAATGCGGCGAGTG 498  
QY 488 GCTGAGTGGGCGAGTGAAGGAGTATGGTCTTCCAGACGCGACATCTTGTGTTCCAGAA 547  
DB 499 GCTGAGTGGGCGTGAAGAAATATGGCTTCCAGAGCTCAACATCTGTTATCCAGAA 558  
QY 548 CATTGATGGGAAGAGTGTGTAAATGACAAAGATGACTTCCAGAGACTCACGCGCAG 607  
DB 559 CATCGATGGGAAGGAATGTGCAAGATGACCAAGACGACTTCCAGAGGCTCACCCCGAG 618  
QY 608 CTATAACCCAGATATCTCTCTGTCACACCTACCTACCTACCTACAGAGA 654  
DB 619 CTAAACCCGACATCTCTCTACATCTCCATCTACCTACCTACAGAGA 665

CB269538 709 bp mRNA linear EST 20-FEB-2003  
LOCUS 108445 Human Fat Cell 5'-Stretch Plus cDNA Library Homo sapiens  
DEFINITION CDNA 5', mRNA sequence.  
ACCESSION CB269538  
VERSION CB269538.1 GI:28444123  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 709)  
AUTHORS Yang,R.-Z.; Shuldiner,A. and Gong,D.-W.  
TITLE EST analysis of human adipose gene expression  
JOURNAL Unpublished (2002)  
COMMENT Contact: Gong Da-Wei  
Division of Endocrinology, Diabetes and Nutrition  
University of Maryland  
660 Redwood St, HH497, Baltimore, MD 21201, USA  
Tel: 410 706 1672  
Fax: 410 706 1622  
Email: dgong@medicine.umaryland.edu  
PCR Primers  
FORWARD: CTCGGAGCGCGCCATCTGTCTGGT  
BACKWARD: ATATGACTACATATAGGCGGAATGG  
Seq primer: GTTGTACCGGGAATC.  
Location/Qualifiers  
1..709  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/sex="Male and Female"  
/tissue\_type="Adipose"  
/clone\_lib="Human Fat Cell 5'-Stretch Plus cDNA Library"  
/note="Vector: lambdaTriplex"

ORIGIN  
Query Match 33.4%; Score 483.4; DB 14; Length 709;  
Best Local Similarity 84.0%; Pred. No. 3.6e-133;  
Matches 593; Conservative 0; Mismatches 107; Indels - 6; Gaps 4;  
QY 667 TTATTTTTCCTCAATACATCAGTTTACCCGAGACGCAACGCAAGAAATACCAAGAGCCAG 726  
DB 6 TTATTTTTCCTCAATACATCAGTTTACCCGAGACGCAACGCAAGAAATACCAAGAGCCAG 65  
QY 727 ATTATCTTATGAGCAAGCGAGGAGATCAGCGTGGACGAGTCCAGAGCTCACAGCCATCCCA---CTC 783  
DB 66 ATTATCTTATGAGCAAGCGAGGAGATCAGCGTGGACGAGTCCAGAGCTCACAGCCATCCCA 125  
QY 784 AGTCAAAAGCTACCAACCATCATCTTCAACAGTGGCCGAAACAGACAGAGCAGCTCCCTC 843  
DB 126 AGTCGAAAGCTGCTCAACCATCTCTTCCACAGTGGCCGAAACAGACAGAGCTCCCTC 185  
QY 844 AGTTAGATCTTATCAGATTTCTTGGACCCAGCAGAGCGCTTGTGCAATCCAGGAGTG 903  
DB 186 AGTTAGATCTTATCAGATTTCTTGGACCAACAAAGTAGCAGCTTGCATAATCCAGGCGAGTG 245  
QY 904 GGCAGATACAGCTATGGCAGTTCCTTACTGAGCTTCTGCGACAGCTCCAACTCCAACT 963  
DB 246 GCCAGATCCAGCTTTGGCAGTTCCTTCTGAGCTCCTGTCGACAGCTCCAACTCCAGCT 305  
QY 964 GCATCACTTGGGAGGCGCACAAATGGGAGTTCAGAGATGACAGACCTGATGAGTGGCTC 1023  
DB 306 GCATCACTTGGGAGTTCACACGCGGAGTTCAGATGACGATCCCGAGAGTGGCC 365  
QY 1024 GCGGTTGGGAGAGAGAGAAAGCAACCTAACATGAACTATGACAAACTCAGCCGTCAC 1083  
DB 366 GCGGCTGGGAGAGCGGAAGAGCAACCCCAACATGAATACGATAGCTCAGCGCGGCC 425  
QY 1084 TTGCTACTACTATGACAAATATTTATGACTAAGTTTATGTTAAAGTTTATGTTAAAGTTTATG 1143  
DB 426 TCGGTTACTACTATGACAAATATTTATGACTAAGTTTATGTTAAAGTTTATGTTAAAGTTTATG 485  
QY 1144 AATTGATTTCCAGCAATCGCTCAGCGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCT 1203  
DB 486 AGTTGCACTTCCAGCGGATCGCCAGCGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCT 545  
QY 1204 ACAATACCATCAGACCTCCCTTACATGATGTTTCTTACATGACACACCCCGAGAGATGA 1263  
DB 546 ACAAGTACCCCTCAGACCT-CCGTATACGTTGGTCTTATCAGCGCCACCCACAGAGATGA 604  
QY 1264 ACTTTGTAGTCTCCCATCCCTGCTTCCGCTTACCTTCCAGCTTCCAGCTTCCAGCTTCCAGCT 1323  
DB 605 ACTTTGTGGCGCCCGCCCTTCCAGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCT 662  
QY 1324 CTATCCATCTGGAATTCACCAACTGGAGGCTTACCCCAATAC 1369  
DB 663 CAAACCCATCTGGAATTCACCAACTGGAGGCTTACCCCAATAC 708

RESULT 15  
EG388025 715 bp mRNA linear EST 12-MAR-2001  
LOCUS 602412867F1 NIH\_MGC\_92 Homo sapiens cDNA clone IMAGE:4521233 5',  
DEFINITION mRNA sequence.  
ACCESSION EG388025  
VERSION EG388025.1 GI:13281471  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 715)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-r@mail.nih.gov

Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LHAM10420 row: f column: 18  
 High quality sequence stop: 693.  
 Location/Qualifiers  
 1. 715  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4521233"  
 /tissue\_type="embryonal carcinoma, cell line"  
 /lab\_hosts="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_92"  
 /note="Organ: testis; Vector: pCMV-SPORT6; Site 1: NotI;  
 Site 2: SalI; Cloned unidirectionally; oligo-dT primed.  
 Average insert size 2.5 kb. Library enriched for  
 full-length clones and constructed by Life Technologies.  
 Note: this is a NIH\_MGC Library."

# FEATURES

source

## ORIGIN

Query Match	33.3%	Score 482.2;	DB 12;	Length 715;
Best Local Similarity	84.1%	Pred. No. 8.2e-133;		
Matches	544;	Conservative	0;	Mismatches 103; Indels 0; Gaps 0;

  

Qy	8	CGCAACGATTAATTTATTATTAGCAATTTATTAGCATCAATAATCTTGATCACATTATGGC	67
Db	39	GAGAAGCAATTAATTTATTAAACATTTATTAAACATCAATAAATGATTGCATTATGGC	98
Qy	68	AAGCACTATTAAAGGAAGCATTTATCAGTGGTGAAGACCAAGCCAGTCCTTGTGAGTGTGC	127
Db	99	CAGCACTATTAAAGNAGCCITTATCAGTTGTGAGTGAGGACCAAGTCGTTGTTGAGTGTGC	158
Qy	128	CTACGGATCGCCCAACCTTGCAAGACAGAAATGACAGCCCTCCTCCAGTGAATATGG	187
Db	159	CTACGGAAACGCCACACCTGGCTAAGACAGAGATGACCGCGTCCTCCTCCAGCGACTATGG	218
Qy	188	GCMAACATCAAGATGAGCCCGCGTTCCCGCAGGAGCTGTTATCAGAGCCCGCCGGC	247
Db	219	ACAGACTTCCAGATGAGCCACCGGTCCTCTCAGCAGATTTGGTGTCTCAACCCCGAGC	278
Qy	248	CAGAGTTACCATTAAGATGAGTGTAAACCCAAACAGGTTAATGGGTCAAGGAATTCACC	307
Db	279	CAGGTCACCATCAAAATGGAATGTAACTTAGCCAGGTGAATGGCTCAGGAACTCTCC	338
Qy	308	TGATGACTGCAGCGTGGCAAGAGGAGAAATGGTTAGCAGTTCAGACATGTTGGAT	367
Db	339	TGATGAATGCAGTGTGGCCAAAGCGGGAAGATGGTGGCAGCCCGACACCGGTGGAT	398
Qy	368	GAACTATGGAAGCTACATGGAAGAGAGCATATTCGGCTCCAAATATGACAAACAATGA	427
Db	399	GAACTACGGCAGCTACATGGAGGAGAGACATGCCACCCCAACATGACCAAGACGA	458
Qy	428	ACGAAGAGTTATTGTCCAGCAGATCCTACGTTATGGAGCAGACAGCATGTACGGCAGTG	487
Db	459	GCCCAAGATTATCGTGCCAGCAGATCCTACGCTATGGAGTACAGACCATGTGCGGCGAGTG	518
Qy	488	GCTGAGTGGGCGAGTGAAGGAGTATGGTCTTCCAGACGTGGACATCTTTGTTTCCAGAA	547
Db	519	GCTGAGTGGGCGGTGAAGATATGGCTTCCAGACGTCAACATCTTTGTTTCCAGAA	578
Qy	548	CATTGATGGGAAGAGTTGTGTAAATGACCAAAGATGACTTCCAGACTCAGCCGAG	607
Db	579	CATCGATGGGAAGAACTGTGCAAGATGACCAAGGACGACTTCCAGAGGCTCAACCCOAG	638
Qy	608	CTATAACGAGATATCCTCCTGTCTACACCTACCTACCTCAGAGAGA	654
Db	639	CTACACGCGGACATCCTCTCTCACATCTCAATACCTCAGAGAGA	685

Search completed: May 27, 2004, 17:47:59  
 Job time : 4104 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 27, 2004, 19:27:37 ; Search time 5617 Seconds  
(without alignments)  
3480.098 Million cell updates/sec

Title: US-09-902-772-2

Perfect score: 2444

Sequence: 1 MASTIKALSVDQSLFE.....IYENTRLPAAHMPSHLGTYV 451

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2.1/USPTO.spool/US0902772/runat\_27052004.162841.2144/app\_query.fasta\_1.647  
-DB=GenEmbl -QMT=fastcap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=500 -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US0902772 @CGN 1.1 3731 @runat 27052004.162841.2144 -NCPV=6 -ICPV=3  
-NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl :  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.scs.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.ro.\*  
26: em.sts.\*  
27: em.un.\*  
28: em.un.\*

29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rtd.\*  
36: em.htg.mam.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match %	Query Length	DB ID	Description
1	2444	100.0	1413	5	AY065661 Gallus ga
2	2444	100.0	1447	6	E31253 Protein hav
3	2414.5	98.8	1516	5	X77159 G.gallus ER
4	2414.5	98.8	1528	6	E31254 Protein hav
5	2327.5	95.2	1509	6	AX657228 Sequence
6	2313	94.6	2016	9	BC040168 Homo sapi
7	2287	92.8	2209	10	AB073080 Mus muscu
8	2192	89.7	2012	5	XLAJ4126 Xenopus l
9	2182	89.3	3166	9	M17254 Human erg2
10	2179	89.2	1808	10	AB031088 Rattus no
11	2159.5	88.4	2133	10	AB073079 Mus muscu
12	2153.5	88.1	1798	5	XLAJ4125 Xenopus l
13	2151	88.0	2131	10	AB073078 Mus muscu
14	1955.5	80.0	1391	6	AX657230 Sequence
15	1946	78.8	1372	6	AX657234 Sequence
16	1840	74.1	3126	9	HUMERG11
17	1807.5	74.0	1219	6	AX657232 Sequence
18	1637	67.0	1436	5	DRE249590 Danio rer
19	1624.5	66.5	1359	6	AX695583 Sequence
20	1624.5	66.5	1359	9	AY029368 Homo sapi
21	1624.5	66.5	2394	9	BC010115 Homo sapi
22	1624.5	66.5	2316	6	A36461 Sequence 2
23	1624.5	66.5	2398	6	AR080101 Sequence
24	1624.5	66.5	2398	9	X67001 H.sapiens H
25	1624.5	66.5	2957	6	AX587825 Sequence
26	1624.5	66.5	2957	6	AX695582 Sequence
27	1624.5	66.5	2957	9	HUMERGBFLI
28	1624.5	66.5	2959	9	BC001670
29	1624.5	66.5	6849	12	AY029367 Synthetic
30	1621.5	66.3	1673	9	S45205
31	1613.5	66.0	1359	6	AX695580 Sequence
32	1613.5	66.0	1729	6	AX695579 Sequence
33	1613.5	66.0	1729	10	MMFLI1
34	1607	65.8	2919	5	BC056588 Danio rer
35	1599	65.4	2854	6	BD094081 Shear str
36	1579.5	64.6	3490	5	CCRNAPLI
37	1576	64.6	4403	5	XLFLIG
38	1567.5	64.1	3545	5	CCFLIONCO
39	1536.5	62.9	1932	9	HUMFLI1
40	1486.5	60.8	1245	5	AF177538 Danio rer
41	1245	51.0	2722	5	BC055627
42	1227	50.2	816	10	S66169
43	1183.5	48.4	762	9	S72821
44	965	39.5	1431	9	AF327066
45	956.5	39.1	14518	9	AP001423 Homo sapi

# ALIGNMENTS

RESULT 1

AY065661  
 LOCUS 1413 bp mRNA linear VRT 17-DEC-2001  
 DEFINITION Gallus gallus erg isoform C-1-1 mRNA, complete cds; alternatively spliced.  
 ACCESSION AY065661  
 VERSION AY065661.1 GI:17887440  
 KEYWORDS  
 SOURCE Gallus gallus (chicken)  
 ORGANISM Gallus gallus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.  
 1 (bases 1 to 1413)  
 Iwamoto, M., Higuchi, Y., Koyama, E., Enomoto-Iwamoto, M., Kurisu, K., Yeh, H., Rosenbloom, J. and Pacifici, M.  
 Transcription factor ERG variants and functional diversification of chondrocytes during limb long bone development  
 J. Cell Biol. 150 (1), 27-40 (2000)  
 20351415  
 10893254  
 10893254  
 2 (bases 1 to 1413)  
 Iwamoto, M., Higuchi, Y., Enomoto-Iwamoto, M., Kurisu, K., Koyama, E., Yeh, H., Rosenbloom, J. and Pacifici, M.  
 The role of ERG (ets related gene) in cartilage development  
 Osteoarthritis. Cartil. 9 Suppl A, S41-S47 (2001)  
 21535378  
 11680687  
 3 (bases 1 to 1413)  
 Iwamoto, M., Higuchi, Y., Yeh, H. and Pacifici, M.  
 Direct Submission  
 Submitted (05-DEC-2001) Oral Anatomy and Developmental Biology, Osaka University Faculty of Dentistry, 1-8 Yamadaoka, Suita, Osaka 565-0871, Japan  
 Location/Qualifiers  
 1..1413  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9031"  
 37..11362  
 /note="alternatively spliced; transcription factor"  
 /codon\_start=1  
 /product="erg isoform C-1-1"  
 /protein\_id="AA140889.1"  
 /db\_xref="GI:17887441"  
 /translation="VASTIKELSVSEDSQIFBCAYGSPHLAKTEMTASSSEYGGT SKSPRVQQWLSQPPARVTIKMECPNQVNGSRNPDCCSVAGKRMVSSSDNVGM NYGSMKFKHPPNPNTTNRVIVPADPTLWSTDRVQWLEWAKVGYGPDVDILF QNDGKELCKTKDQFRLTPSYNADILLSHLYLRERGAFTI FPNSTVYPEATQRT TRPDLPYQARRSAMTSHTQSKATQPSSTVPKTEDQRPQDPYQILGPTSSRLA NPGSQIOLWQFLLLELSDSNSNCITWETNGTBFKMTDPDEVARRWGRKSKPNMY DKLSALRYVYDKMTWKVHGKRYAYKEDFHGIAQLQPHPPBSMYKVPSDLPYMS YAHAPQKNFVAPHPALPVTSSSFPAPNPYNSPTGGIYVPTLELPAHMPHSLGY Y"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 2,02e-162 Length: 1413  
 Score: 2444.00 Matches: 451  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 5 Gaps: 0  
 US-09-902-772-2 (1-451) x AY065661 (1-1413)  
 QY 1 MetAlaSerThrIleisGluAlaLeuSerValSerGluAspGlnSerLeuPheGlu 20  
 DB 37 ATGGCAAGCATTATTAAGAAAGCATATCAGTGGTGAAGACCATCTTGTGTGAG 96  
 QY 21 CysAlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSertGlu 40  
 DB 97 TGTGCTACGGATCGCCCACTTCGACAGACAGAAATGACAGCTCTCTCTCCAGTGA 156

QY 41 TyrGlyGlnThrSerIysMetSerProArgValProGlnGlnAspTrpLeuSerGlnPro 60  
 DB 157 TATGGCAAAATCAAGATGAGCCGCGCTTCCCAGCAGACTGGTTATCAGAGCC 216  
 QY 61 ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80  
 DB 217 CCGGCCAGAGTTACCATTAAGATGGAGTGTAAACCCAAACACAGGTTAATGGGTCAAGGAAT 276  
 QY 81 SerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsnVal 100  
 DB 277 TACCTGATGATCGACGCTGCAANAGGAGGAATAATGTTAGCAGTTACAGCATGTT 336  
 QY 101 GlyMetAsnTyrGlySerTyrMetGluLysHisIleProProProProAsnMetThrThr 120  
 DB 337 GGGATCAACTATGGAAGCTACATGGAAGAGAGACATATTCGCCCTCCAAATATGACAACC 396  
 QY 121 AsnGluArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArg 140  
 DB 397 AATGAACGAGAGTATTATGTGCAGCAGATCTCTAGTTAAGGACACAGCATGTACGG 456  
 QY 141 GlnTrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPhe 160  
 DB 457 CAGTGGCTGGAGTGGGACGTGAAGGAGTATGTCTTCCAGACGTGGACATCTTGTGTTC 516  
 QY 161 GlnAsnIleAspGlyLysGluLysMetThrLysAspAspPheGlnArgLeuThr 180  
 DB 517 CAGAACATTGATGGGAAGAGTGTGTAAATGACCAAGATGACTTCCAGAGACTCAG 576  
 QY 181 ProSerTyrAsnAlaAspIleLeuSerHisLeuHisTyrLeuArgGluArgGlyAla 200  
 DB 577 CCGAGCTATACCGACATATCTCTGTCCACACCTACACTACTCTCAGAGAGAGAGGCC 636  
 QY 201 ThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrArg 220  
 DB 637 ACTTTTATTTTCCAAATACATCAGTTTACCAGAGCAACCAAGATACCAACAAAG 696  
 QY 221 ProAspLeuProTyrGluGlnAlaArgArgSerAlaTrpThrSerHisProThr 240  
 DB 697 CCGATTATCTTATGCAAGCAGGAGATCAGCGTGGACGAGTCCACAGCCATCCACT 756  
 QY 241 GlnSerLysAlaThrGlnProSerSerSerThrValProLysThrGluAspGlnArgPro 260  
 DB 757 CAGTCAAAAGCTACCCCAACCATCATCTTCAACAGATGTCCTCAACCAAGACCAAGAC 816  
 QY 261 GlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGlySer 280  
 DB 817 CAGTTAGATCCTTATCAGATTCTTGGACCGACAGCGCCGCTTTCCTCAATCCAGGAGT 876  
 QY 281 GlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeuLeuSerAspSerSerAsnSerAsn 300  
 DB 877 GGGCAGATACAGCTATGGCAGTCTCTACTGGAGCTTCTGTCGACAGCTCCCACTCAAC 936  
 QY 301 CysIleThrTrpGluGlyThrIleGlyGluPheLysMetThrAspProAspGluValAla 320  
 DB 937 TGCATCACCTGGAGGGGCAAAATGGGAGTTCAAGATCAGACACCTCTGATGAGTGGCT 996  
 QY 321 ArgArgTrpGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArgAla 340  
 DB 997 CGCGCTGGGAGAGAGAGAAAGCAACCTTAACATGAACTATGACAAACTCAGCCGTGCA 1056  
 QY 341 LeuArgTyrTyrThrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaTyr 360  
 DB 1057 CTTCCGTACTACTATGACAAAATATATGACTAAAGTTGATGGTAAACGCTATGCTTAC 1116  
 QY 361 LysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProProGluSerSerMet 380  
 DB 1117 AANTTTGATTTCCACGGAATCGCTCAGGCCCTCCAGCCCTCACCATCATCATG 1176  
 QY 381 TyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisProGlnLysMet 400  
 DB 1177 TACAAATACCATCAGACCTCCCTCATGAGTTCCTACCATGCACACCCCGAGAGATG 1236  
 QY 401 AsnPheValAlaAlaProHisProProAlaLeuProValThrSerSerSerPheAlaAla 420



```

Db      1237 AACTTGTAGCTCCCATCCCTGCTTTGGCCGTAACTCATCCAGCTTTTGTCTGCC 1296
QY      421 ProAsnProTyrTrpAsnSerProThrGlyGlyLeuTyrProAsnThrArgLeuProAla 440
Db      1297 CCTAATCATCTAGTGAATTCACCAACTGGAGGCACTTACCCCAATACAGGCTCCAGCT 1356
QY      441 AlaHisMetProSerHisLeuGlyThrTyrTyr 451
Db      1357 GCTCATATGCTTCCCATCTTGGCACCCTACTACTAC 1389

RESULT 2
E31253
LOCUS      1447 bp      DNA      linear      PAT 18-JUN-2001
DEFINITION      Protein having cell calcifying inhibitory activity and gene
                  encoding the same.
ACCESSION      E31253
VERSION      E31253.1 GI:13025685
KEYWORDS      JP 199075871-A/1.
SOURCE      unidentified
ORGANISM      unclassified.
REFERENCE      1 (bases 1 to 1447)
AUTHORS      Hiroyasu, I., Yoshinobu, H., Marjio, P., Joel, R. and Helena, E.
TITLE      Protein having cell calcifying inhibitory activity and gene
                  encoding the same
JOURNAL      Patent: JP 199075871-A 1 23-MAR-1999;
                  CHUGAI PHARMACEUT CO LTD, HE TRUSTEES OF THE UNIVERSITY OF
                  PENNSYLVANIA
COMMENT      OS      Unidentified
                  PN      JP 199075871-A/1
                  PD      23-MAR-1999
                  PF      29-MAY-1998 JP 1998166076
                  PR      18-JUN-1997 US 08/878177, 20-JUN-1997 US 60/050297 PI
                  HIROYASU IWAMOTO, YOSHINOBU HIGUCHI, MARIUJO PASHIFIKI, PI JOEL
                  ROZENBUROOM,
                  PI      HELENA E
                  PC      C12N15/09, A61K48/00, C07K14/465, C07K16/18, C12Q1/68 //A61K38/00,
                  PC      C12P21/02,
                  PC      C12P21/08, C12N15/00, A61K37/02
                  CC      Strandedness: Single;
                  CC      Topology: Linear;
                  FH      Key      Location/Qualifiers
                  FT      source      1..1447
                  FT      location/qualifiers      /organism='Unidentified'.
FEATURES
    source
    1..1447
    /organism="unidentified"
    /mol_type="genomic DNA"
    /db_xref="taxon:32644"

ORIGIN

Alignment Scores:
Pred. No.:      2,08e-162      Length:      1447
Score:      2444.00      Matches:      451
Percent Similarity:      100.00%      Conservative:      0
Best Local Similarity:      100.00%      Mismatches:      0
Query Match:      100.00%      Indels:      0
DB:      6      Gaps:      0

US-09-902-772-2 (1-451) x E31253 (1-1447)

QY      1 MetAlaSerThrIleLysGluAlaLeuSerValSerGluAspGlnSerLeuPheGlu 20
Db      63 ATGGCAAGCACTATTAAAGGAAGCATTTATCAGTGGTGAGTGAAGACCACTCTTGTGTGAG 122
QY      21 CysAlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGlu 40
Db      123 TGTGCTCATCGGATCGCGCCCACTTGGAAAGACAGAAATGACAGCCTCTCTTCCAGTGAA 182
QY      41 TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnPro 60
Db      183 TATGGCAACAACATCAAGATGAGCGCGCTTCCCGACGAGCACTGGTTATCAAGGCC 242

```

```

QY      61 ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
Db      243 CCGGCCAGAGTTTACCATTAAGATGGAGTGTAAACCAACCCAGGTTAATGGTCAAGGAAT 302
QY      81 SerProAspAspCysSerValAlaLysGlyGlyLysMetValSerSerSerSerAspAsnVal 100
Db      303 TCACCTGTAGTACCTGCGAGCGTGCAGAAAGAGGAGGAAATGTTAGCAGTTTCAGACCAATGT 362
QY      101 GlyMetAsnTyrGlySerTyrMetGluGluLysHisIleProProProAsnMetThrThr 120
Db      363 GGGATGAACATATGGAAGCTACATGGNAGAGAGCATATTCGCGCTCCAAATATGACAACC 422
QY      121 AsnGluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArg 140
Db      423 AATGAACGAAGAGATTATTGTGCAGCAGATCTCTACGTTTATGGAGCAGACAGACCATGTACCG 482
QY      141 GlnTrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeuPhe 160
Db      483 CAGTGGCTGGAGTGGCAGTGAAGAGTATGGTCTTCCAGACGTGGACATCTTGTGTTC 542
QY      161 GlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThr 180
Db      543 CAGAACATTGATGGGAAGAGATTGTCTAAATACCAAAAGATGACTTCCAGAGACTCACG 602
QY      181 ProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArgGlyVala 200
Db      603 CCGAGCTATACGAGATATCTCTCTGTCACCTACACTACTCTCAGAGAGAGAGAGGCC 662
QY      201 ThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrArg 220
Db      663 ACTTTTATTTTCCAAATACATCAGTTTACCCAGAGAACCGCAAGAATAACACAAGG 722
QY      221 ProAspLeuProTyrGluGlnAlaArgArgSerAlaTrpThrSerHisSerHisProThr 240
Db      723 CCAGATTACCTTATGAGCAAGGAGAGATCAGCTGAGCAGTCCAGGCTCAGGCCATCCACT 782
QY      241 GlnSerLysAlaThrGlnProSerSerSerThrValProLysThrGluAspGlnArgPro 260
Db      783 CAGTCAAAAGCTACCCCAACCATCATCTTCAACAGTGCCTCCCAAAACAGAGACCAAGCTCCT 842
QY      261 GlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGlySer 280
Db      843 CAGTTAGATCCTTATCAGATTCTTGGACCGACAGCAGCGCTCTTCAAAATCCAGGAGT 902
QY      281 GlyGlnIleGlnLeuTrpGlnPheLeuLeuLeuSerAspSerSerAsnSerAsn 300
Db      903 GGGCAGATACAGCTATGGCAGTTCCTACTGGAGCTTCTGTCGGACAGCTCCACTCCAAC 962
QY      301 CysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspProAspGluValAla 320
Db      963 TGCATCACCTGGAGGGGCACAAATGGGGAGTTCAAGATGACAGACCCCTGATGAAGTGGCT 1022
QY      321 ArgArgTrpGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArgAla 340
Db      1023 CGGCGTTGGGAGAGAGAGAAACCAACCTAATGAACTATGCAAACTCAGACCCGTGCA 1082
QY      341 LeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaTyr 360
Db      1083 CTTTCGCTACTACTATGCAAAAATATATGACTTAAAGTTTCATGTAACCGCTATGCTTAC 1142
QY      361 LysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGluSerSerMet 380
Db      1143 AAATTTGATTTCACCGAATCGCTCGAGCCCTCCAGCCTCACCCCTCAGAAATCATCATG 1202
QY      381 TyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisProGlnLysMet 400
Db      1203 TACAAATACCCATCAGACCTCCCTCATGAGTTCTTACCATGCACACCCCGAAGATG 1262
QY      401 AsnPheValAlaLysProHisProAlaLeuProValThrSerSerPhePheAlaAla 420
Db      1263 AACTTTGTAGTCCCATCCCTCTGCTTTGCCGTAACCTCATCCAGCTTTTGTCTGCC 1322

```

```

QY 421 ProAenProTyrTtpAenSerProThrGlyGlyIleTyrProAenThrArgLeuProAla 440
DB 1323 CCTAATCATCACTGGAATTCACCAACTGGAGGCATCTACCCCAATACAGGCTGCCAGCT 1382
QY 441 AlaHisMetProSerHisLeuGlyThrTyrTyr 451
DB 1393 GCTCATATGCTTCCCATCTTGGCACCTACTAC 1415

RESULT 3
GGERG
LOCUS GGERG 1516 bp mRNA linear VRT 27-APR-1995
DEFINITION G.gallus ERG mRNA.
ACCESSION X77159
VERSION X77159.1 GI:790439
KEYWORDS ERG gene.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1
AUTHORS Dhordain,P., Dewitte,F., Desbiens,X., Stehelin,D. and
Duterque-Coquillaud,M.
TITLE Mesodermal expression of the chicken erg gene associated with
precartilaginous condensation and cartilage differentiation
JOURNAL Mech. Dev. 50 (1), 17-28 (1995)
MEDLINE 95329425
PUBMED 7605748
REFERENCE 2 (bases 1 to 1516)
AUTHORS Duterque-Coquillaud,M.
TITLE Direct Submission
JOURNAL Submitted (17-JAN-1994) M. Duterque-Coquillaud, CNRS UA 1160,
Oncologie Moleculaire, Institut Pasteur, 1 rue Calmette, 59019
Lille, FRANCE
FEATURES
Location/Qualifiers
1..1516
/organism="Gallus gallus"
/mol_type="mRNA"
/db_xref="taxon:9031"
/tissue_type="spleen"
/clone_lib="lambda gt10"
/dev_stage="adult"
1..1516
/gene="ERG"
63..1499
/gene="ERG"
/codon_start=1
/protein_id="CAA54404.1"
/db_xref="GI:790440"
/db_xref="GOA:Q90837"
/db_xref="SWISS-PROT:Q90837"
/translaton="MASTIKELSVSDESQSLFECAYGSPHLAKTMTASSSSEYQQT
SKMSPRVQDQWLSQPPARVTIKMECNFNQVGRNSPDDCSVAKGKXVSSDNVGM
NYGSMYMEKHIPPNMTTNERRTVPADPTLWSTDRVOWLEWAKVEYGLPDVDILF
QNIQDKLCKMTDFFDLTPSYNADILLSHLVRLTLPPLHTSDDDVKALQNSPRL
MHARTGGAFFIPNTSVPEAQRIITRPDLPEQARRSAWTSHPSTQSKATQPS8
STFKPTDQRPDLQPLGFTSRLANPQSGQQLMQFLLELSSNSNCITWEGT
NGVEFPDDEVARWGRKSKNMNYDKLSRLRYDDKNTKMTKVGKRYAKFDFH
GIAQLQHPHPSESMYKPSDLPTMSSYHAHPQKMFVAPHPPALPVTSSSFAEPNP
YWNSTGGIYVNTLRLPAHMFPSHLGTY"

ORIGIN
Alignment Scores:
Pred. No.: 2,57e-160 Length: 1516
Score: 2414.50 Matches: 450
Percent Similarity: 94.14% Conservative: 0
Best Local Similarity: 94.14% Mismatches: 1
Query Match: 98.79% Indels: 27
DB: 5 Gaps: 1

US-09-902-772-2 (1-451) x GGERG (1-1516)
QY 1 MetAlaSerThrIleLysGluAlaLeuSerValSerGluAspGlnSerLeuPheGlu 20

```

```

DB 63 ATGGCAAGCACATATTAAAGGAAGCATATCAGTGGTGAAGACAGCTCTCTGTGTTGAG 122
QY 21 CysAlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGlu 40
DB 123 TGTGCTACGGATCGCCACCTTCGAAGACAGAAATGACAGCCTCTCTTCAGTGA 182
QY 41 TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnPro 60
DB 183 TATGGCGCAACATCAAGATGAGCCCGCGCTTCCCGACGAGTGGTATTACAGGCC 242
QY 61 ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
DB 243 CCGGCCAGAGTTACCATTAAAGATGGAGTGAACCAACACAGGTTAATGGGTCAAGGAAT 302
QY 81 SerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsnVal 100
DB 303 TCACCTGATGAGTCGAGCGCTGGCAAGAGAGGAAATGGTTAGCAGTTTCAGACAATGTT 362
QY 101 GlyMetAsnTyrGlySerTyrMetGluGluLysHisIleProProProAsnMetThrThr 120
DB 363 GGGATGAATATGGAAAGCTCATGGAAGAAAGCATATTCGCCCTCCAAATATGACACCC 422
QY 121 AsnGluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArg 140
DB 423 AATGAACAGAGAGTTATTGTGCCAGCAGATCCTACGTTATGGAGCAGACCATGTACGG 482
QY 141 GlnTrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeuPhe 160
DB 483 CAGTGGCTGGAGTGGCGAGTGAAGGAGTATGCTCTCCAGACGTCGAGCATCTTGTGTTC 542
QY 161 GlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThr 180
DB 543 CAGAACATTTGATGGGAAAGAGTTGTGTAATATGACCAAGATGACTTCCAGAGACTCAG 602
QY 181 ProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArg----- 198
DB 603 CCGAGCTATAACGCGAGATATCCTCTGTCCACACCTACACTACCTCAGAGAGACTCCTCTT 662
QY 198 ----- 198
DB 663 CCACATTTGACTTCAGATGATGTTGATAAGGCCCTTACAAACTCTCCACGGTTAATGCAT 722
QY 199 -----GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAla 213
DB 723 GCTAGAAAACACAGGAGAGGACCATTTTATTTTCCAAATACATCATCATGTTTACCCAGAGCA 782
QY 214 ThrGlnArgIleThrThrArgProAspLeuProTyrGluGlnAlaArgSerAlaTrp 233
DB 783 ACGCAAGAATATACAAACAGGCCAGATTTACCTTATGAGCAAGCGAGAGATCAGCGTGG 842
QY 234 ThrSerHisSerHisProThrGlnSerLysAlaThrGlnProSerSerSerThrValPro 253
DB 843 ACGAGTCAAGCCATCCCACTCAGTCAAAAGCTACCAACCATCATCTTTCAACAGTGCCC 902
QY 254 LysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSer 273
DB 903 AAAACAGAGAACCCAGCGCTCCTCAGTATGATCCTTATCAGATCTTTGGACCCAGCAGCAGC 962
QY 274 ArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuLeuLeu 293
DB 963 CGTCTTGCAAAATCCAGGAGTGGGCAGATACAGCTATGGCAGTTCTCTACTGAGGCTCTG 1022
QY 294 SerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMet 313
DB 1023 TCGGACAGCTCCCACTCCAACTGCATCCTCGGAGGCGCAAAATGGGAGTTCAGATG 1082
QY 314 ThrAspProAspGluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMetAsn 333
DB 1083 ACAGACCTGTGTAAGTGGCTCGGCTGGGAGAGAGGAAAAGCAACCTAACATGAAC 1142
QY 334 TyrAspLysLeuSerArgAlaLeuArgTyrTyrAspLysAsnIleMetThrLysVal 353

```

Db 1143 TATGACAAACTCAGCCGTCGCTACTACTATGACAAAAATATATGACTAAGTT 1202

Qy 354 HisGlyLysAsgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnPro 373

Db 1203 CATGTTAAACGCTATGCTCAAAATTTGATTTCCACGAATCGCTCAGGCCCTCAGCCT 1262

Qy 374 HisProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyr 393

Db 1263 CACCTCCAGATCATCATGACAAATCCATCAGACCTCCCTCATGAGTTCTTAC 1322

Qy 394 HisAlaHisProGlnLysMetAsnPheValAlaProHisProAlaLeuProValThr 413

Db 1323 CATGCACACCCCAAGAGAACTTTGTAGTCCCATCCCATCCCTGCTTCCCGCTAAC 1382

Qy 414 SerSerSerPhePheAlaProAsnProTyrTyrAsnSerProThrGlyLysTyr 433

Db 1383 TCATCCAGCTTTTGTGCCCCCTAATCCATGCTGGAATCCACAACTGGAGGATCTTAC 1442

Qy 434 ProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451

Db 1443 CCCATACAGGCTCCAGCTGCTCATATGCTTCCCATCTTGGCACCTACTAC 1496

RESULT 4

E31254

LOCUS

DEFINITION

Protein having cell calcifying inhibitory activity and gene

encoding the same.

ACCESSION

E31254

VERSION

E31254.1

KEYWORDS

GI:13025686

SOURCE

JP 1999075871-A/2.

ORGANISM

unidentified

REFERENCE

1 (bases 1 to 1528)

AUTHORS

Hiroyasu, I., Yoshinobu, H., Marijo, P., Joel, R., and Helena, E.

TITLE

Protein having cell calcifying inhibitory activity and gene

JOURNAL

encoding the same

Patent: JP 1999075871-A 2 23-MAR-1999;

CHUGAI PHARMACEUT CO LTD, HE TRUSTEES OF THE UNIVERSITY OF

PENNSYLVANIA

COMMENT

OS Unidentified

PN JP 1999075871-A/2

PD 23-MAR-1999

PF 23-MAY-1998 JP 1998166076

PR 18-JUN-1997 US 08/978177, 20-JUN-1997 US 60/050297 PI

HIROYASU IWAMOTO, YOSHINOBU HIGUCHI, MARIJO PASHIFIKI, PI JOEL

ROZENBUROOM,

PI HELENA E

PC C12N15/09, A61K48/00, C07K14/465, C07K16/18, C12Q1/68//A61K38/00,

PC C12P21/02, C12N15/00, A61K37/02

PC C12P21/08, C12N15/00, A61K37/02

CC Strandedness: Single;

CC Topology: Linear;

FH Key

FT Location/Qualifiers

FT source 1..1528

FT /organism='Unidentified'.

FT Location/Qualifiers

1..1528

/organism='unidentified'

/mol\_type='genomic DNA'

/db\_xref='taxon:32644'

ORIGIN

Alignment Scores:

Pred. No.: 2,6e-160 Length: 1528

Score: 2414.50 Matches: 450

Percent Similarity: 94.14% Conservative: 0

Best Local Similarity: 94.14% Mismatches: 1

Query Match: 98.79% Indels: 27

DB: 6 Gaps: 1

US-09-902-772-2 (1-451) x E31254 (1-1528)

Qy 1 MetAlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20

Db 63 ATGGCAAGCAGCTATTAAAGAAAGCATTATCAGTGGTGAAGTGAAGCAGCCTCTTGTGAG 122

Qy 21 CysAlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGlu 40

Db 123 TGTGCTACGGATCGCCACCTTGCAAAAGACAGAAATGACAGCCTCTCTCCAGTGAA 182

Qy 41 TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspThrLeuSerGlnPro 60

Db 183 TATGGCAACATCAAGATGAGCCGCGCTTCCCAAGCAGGACTGGTATCACAGCCC 242

Qy 61 ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80

Db 243 CCGCCAGAGATTACCATTAAGATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 302

Qy 81 SerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerSerAspVal 100

Db 303 TCACCTCATGACTGCGAGCGTGGCAAAAGGAGGAAAATGTTAGCAGTTTACAGCAATGT 362

Qy 101 GlyMetAsnTyrGlySerTyrMetGluGluLysHisLeuProProProAsnMetThrThr 120

Db 363 GGGATGAATATGGAAGCTACATGGAAGAGAGCATATTCGCGCTCCAAATATGACAACC 422

Qy 121 AsnGluArgArgValIleValProAlaAspProThrLeuThrSerThrAspHisValArg 140

Db 423 AATGAACGAAGAGTATTGTGCCAGCAGATCCTTACGTTTATGAGCAGCAGACCATGACGG 482

Qy 141 GlnThrLeuGluThrAlaValLysGluTyrGlyLeuProAspValAspLysLeuLeuPhe 160

Db 483 CAGTGGCTGGAGTGGCAGTGAAGAGTATGGTCTTCCAGAGCTGGACATCTTGTGTTC 542

Qy 161 GlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThr 180

Db 543 CAGAACATTGATGGGAAAGAGTTGTGTAATAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 602

Qy 181 ProSerTyrAsnAlaAspIleLeuSerHisLeuHisLeuArgGluArg 198

Db 603 CCGAGCTATAACGAGATATCTCTCTGTACACCTACATCTCTCAGAGAGACTCTCTCT 662

Qy 198 ----- 198

Db 663 CCACATTTCAGCTTCAGATGATGTTGATAAGGCTTACAAAACCTCTCCACGGTTAATGCAT 722

Qy 199 -----GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAla 213

Db 723 GCTAGAAAACACAGGAGGAGCCACTTTTATTTTCCAAATACATCAGTTTACCAGAGCA 782

Qy 214 ThrGlnArgIleThrThrArgProAspLeuProTyrGluGlnAlaArgSerAlaTyr 233

Db 783 ACGCAAGATATACAAACAGGCGCAGATTTACCTTATGACCAAGCAGGAGATCAGCGTGG 842

Qy 234 ThrSerHisSerHisProThrGlnSerLysAlaThrGlnProSerSerSerThrValPro 253

Db 843 ACGAGTCACAGCCATCCCACTCAGTCAAAAGTGTACCAACCATCATCTTCAACAGTGCCT 902

Qy 254 LysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSer 273

Db 903 AAAACAGAGACAGCGCTCTCAGTTAGATCTCTTATCAGATCTTTGACCGCAGCAGCAG 962

Qy 274 ArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuThrGlnPheLeuLeuLeuLeu 293

Db 963 CGTCTTCAAAATCCAGGAGTGGCAGATACAGTATGCGAGTCTCTACTGGAGCTTCTG 1022

Qy 294 SerAspSerSerAsnSerAsnSerLysIleThrTyrGluGlyThrAsnGlyGluPheLysMet 313

Db 1023 TCGGACAGCTCCCACTCCCACTGACATCAGCTGGAGGGGACAAATGGGGAGTTCAAGATG 1082

Qy 314 ThrAspProAspGluValAlaArgArgTyrGlyGluArgLysSerLysProAsnMetAsn 333

Db 1083 ACAGACCTGATGAAGTGGCTCGCGTGGGAGAGAGGAAAGCAAACTAATCAATGAAC 1142

Qy 334 TyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysVal 353

Db 1143 TATGACAACTCAGCGTGCATCTCGCTACTACTATGACAAAATATATGACTAAAGTT 1202  
 Qy 354 HisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnPro 373  
 Db 1203 CATGGTAAACGCTATGCTACAAATTTGATTCACCGAATCGCTCAGGCGCTCAGGCT 1262  
 Qy 374 HisProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyr 393  
 Db 1263 CACCGCTCAGAAATCATCATGATACAAATACCATCAGAGCTCCCTCATCATGAGTTCTTAC 1322  
 Qy 394 HisAlaHisProGlnLysMetAspPheValAlaProHisProAlaLeuProValThr 413  
 Db 1323 CATGCACACCCCGAGAGATGAACTTTGTAGCTCCCATCCCTGCTTTGCCGTAACC 1382  
 Qy 414 SerSerPhePheAlaAlaProAsnProTyrTyrPheAsnSerProThrGlyGlyIleTyr 433  
 Db 1383 TCATCCAGCTTTTGTGCTGCCCTTAATCCATCTGGAATTCACCAATCGGAGGATCTAC 1442  
 Qy 434 ProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451  
 Db 1443 CCCAATACAGGCTGCCAGCTGCTCATATGCTTCCCATCTTGGCACCTACTAC 1496

RESULT 5  
 AX657228  
 LOCUS AX657228 1509 bp DNA linear PAT 22-MAR-2003  
 DEFINITION Sequence 1 from Patent WO03000724.  
 ACCESSION AX657228  
 VERSION AX657228.1 GI:29160006  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1  
 Luyten, F., de Bari, C. and Dell'Accio, F.  
 Polynucleotide sequences and vectors useful for the prevention or  
 treatment of bone- or cartilage-related disorders  
 Patent: WO 03000724-A 1 03-JAN-2003;  
 Tigenix N.V. (BE)

FEATURES  
 source  
 1..1509  
 /organism="Homo sapiens"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:9606"  
 67..1446  
 /note="unnamed protein product"  
 /codon\_start=1  
 /db\_xref="GI:29160007"  
 /protein\_id="CAD79711.1"

CDS  
 67..1446  
 /note="unnamed protein product"  
 /codon\_start=1  
 /db\_xref="GI:29160007"  
 /protein\_id="CAD79711.1"  
 /translation="MIQVDPAAHKEALSVMSEDSQSLFECAYGTTPHLAKTENTASS  
 SSDYQTSKMSRPVQDWLSQPPARVTIKMCPNSQVNGSRNSPDECSVAKGKRWG  
 SPDVTVMYNGSVEBKMPNMTTNERRVIPADPTLWTDHVRQWLEWAVEYGLP  
 ENILFONIDGKELCKMKDDFORLTSPYNADILLSHLYLRERGAFTFPNTSVYP  
 EATQRTTRPLPEPPRRTAWGHPTQSKAQPSPSTVPKTEDQRLQDPQIL  
 GPTSRLANPQSGQLWQFLLELLSDSSNSCITWETNGEFTKMTDPDQVARMGER  
 KSKPNNDYKLSRALRYYYDKNIMTKVHGKRYAFDFHQAQALQHPPESSLYKIP  
 SDLPYMGSHYAHPOKQNFVAPHPALPVTSSFFFAAPNPYNSPTGGIYPTNRLPTSH  
 NPShLGLTYV"

## ORIGIN

Alignment Scores:  
 Pred. No.: 3,248-154 Length: 1509  
 Score: 2327.50 Matches: 427  
 Percent Similarity: 96.90% Conservative: 10  
 Best Local Similarity: 94.68% Mismatches: 13  
 Query Match: 95.23% Indels: 1  
 DB: 6 Gaps: 1

US-09-902-772-2 (1-451) x AX657228 (1-1509)

Qy 2 AlaSerThrIleLysGluAlaLeuSerValSerGluAspGlnSerLeuPheGluCys 21

Db 91 GCAGCTCATATCAAGGAAGCCCTTATCAGTTGTGAGTGAGGACCACTCGTTGTTGAGTGT 150  
 Qy 22 AlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerSerGluTyr 41  
 Db 151 GCCTACGGAACGCCACACTGGCTAAGACAGAGATGACCGGCTCTCTCTCCAGCGACTAT 210  
 Qy 42 GlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnProPro 61  
 Db 211 GGACAGACTTCCAAGATGAGCCCGCTCTCTCAGCAGGATTTGGTCTCTCAACCCCA 270  
 Qy 62 AlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSer 81  
 Db 271 GCCAGGGTCACCATCAAAATGGAATGTAACCTCAGGAGTGAATGGTCAAGGAACTCT 330  
 Qy 82 ProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsnValGly 101  
 Db 331 CCGTGAATGATGAGTGGCCAAAGCGGGAAGATGGTGGGCGAGCCGACACACCGTTGGG 390  
 Qy 102 MetAsnTyrGlySerTyrMetGluGlyLysIleProProProAsnMetThrThrAsn 121  
 Db 391 ATGAACACTACGGCAGCTACATGGAGGAGGACACATGCCCCCAACATGACACGCAAC 450  
 Qy 122 GluArgValIleValProAlaAspProThrLeuThrSerThrAspHisValArgGln 141  
 Db 451 GAGCGCAGAGTTATCTGCGCAGCAGATCTACGCTATGGAGTACAGACCATGTGGCGAG 510  
 Qy 142 TrpLeuGluTyrAlaValLysGlyTyrGlyLeuProAspValAspIleLeuLeuPheGln 161  
 Db 511 TGGCTGGAGTGGCGGTGAAGAAATATGGCTTCCAGACGTCACATCTTTGTTATTCAG 570  
 Qy 162 AsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThrPro 181  
 Db 571 AACATCGATGGGAAGAACTGTGCAAGATGACCAAGGAGACTTCAGAGGCTCACCCCC 630  
 Qy 182 SerTyrAsnAlaAspIleLeuSerHisLeuHisTyrLeuArgGluArgGlyAlaThr 201  
 Db 631 AGCTACAAACGCCGACATCTCTCTCACATCTCCACTACCTCAGAGAGAGGGGTGCAGCT 690  
 Qy 202 PheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrThrArgPro 221  
 Db 691 TTTATTTTCCAAATACTTTCAGTATATCTGAAGCTACGCAAGAAATACAACTAGGCCA 750  
 Qy 222 AspLeuProTyrGluGlnAlaArgArgSerAlaTyrThrSerHisSerHisProThr 240  
 Db 751 GATTTACCATATGAGCCCCCAGGAGATCAGCTGGACCGGTACGCCGCCACCCACGCC 810  
 Qy 241 GlnSerLysAlaThrGlnProSerSerSerThrValProLysThrGluAspGlnArgPro 260  
 Db 811 CAGTCGAAAGCTGCTCAACCATCTCTCTCCACAGTGCCTCCCAAACTGAAGACCGCTCT 870  
 Qy 261 GlnLeuAspProTyrGlnIleLeuGlyProThrSerArgLeuAlaAsnProGlySer 280  
 Db 871 CAGTTAGATCTCTTATCAGATTCTTGACCAACAAGTAGCCGCTTGCAATTCAGGCACT 930  
 Qy 281 GlyGlnIleGlnLeuTyrGlnPheLeuLeuGluLeuSerAspSerSerAsnSerAsn 300  
 Db 931 GGCCAGATCCAGCTTTGGCAGTTCTCTCTGGAGTCTCTGTCGACAGCTCCAACTCCAG 990  
 Qy 301 CysIleThrTyrProGlyThrAsnGlyGluPheLysMetThrAspProAspGluValAla 320  
 Db 991 TGCATCAGCTGGGAAGGCCCAACCGGAGTTCAGATGACGATCCCCAGGAGTGCC 1050  
 Qy 321 ArgArgTyrGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArgAla 340  
 Db 1051 CGGCGCTGGGAGCGGAGAGCAAAACCCACATGACATGATAGCTCAGCCGCGCC 1110  
 Qy 341 LeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaTyr 360  
 Db 1111 CTCGTTACTACTATGCAAGAACATCATGACCAAGTCCATGGGAGCGCTACGCTAC 1170  
 Qy 361 LysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGluSerSerMet 380

Db 1171 AAGTTCGACTTCACGGGATCGCCAGGCGCTCCAGCCGCCACCCCGGAGTCATCTCTG 1230

Qy 381 TyrIysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisProGlnLysMet 400

Db 1231 TACAAGTACCCCTCAGACCTCCCGTACATGGGCTCTATCACGCCACCCACAGAGATG 1290

Qy 401 AsnPheValAlaProHisProAlaLeuProValThrSerSerPhePheAlaAla 420

Db 1291 AACTTTGGGGCCGCCCTCCAGCCCTCCCGGTGACATCTCCAGTTTTCGTGCC 1350

Qy 421 ProAsnProTyrTrpAsnSerProThrGlyGlyIleTyrProAsnThrArgLeuProAla 440

Db 1351 CCAAAACCACACTGGAATTCACCAACTGGGGGTATATACCCCAACACTAGGCTCCCCACC 1410

Qy 441 AlaHisMetProSerHisLeuGlyThrTyrTyr 451

Db 1411 AGCCATATGCTCTCTCCTGCGGCACCTACTAC 1443

RESULT 6

LOCUS BC040168

DEFINITION Homo sapiens v-ets erythroblastosis virus E26 oncogene like (avian), transcript variant 1, mRNA (CDNA clone MGC:49020 IMAGE:6052140), complete cds.

ACCESSION BC040168

VERSION BC040168.1 GI:25304065

KEYWORDS MGC.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 2016)

AUTHORS Klausner R.D., Collins P.S., Wagner L.L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore I., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C., Raha S.C., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay D.J., Hulyk S.W., Villalón D.K., Muzny D.B., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Heiton E., Kettman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shavchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S., Krzywicki M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E., Jones S.J. and Marra M.A.

TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

MEDLINE 22388257

PUBMED 12477932

REFERENCE 2 (bases 1 to 2016)

AUTHORS Strausberg R.

TITLE Direct Submission

JOURNAL Submitted (22-NOV-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk

Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Institute for Systems Biology

<http://www.systemsbio.org>

contact: [amadnan@systemsbiology.org](mailto:amadnan@systemsbiology.org)

Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAK Plate: 84 Row: O Column: 9.

Location/Qualifiers

1..2016

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="MGC:49020 IMAGE:6052140"

/tissue\_type="Testis, embryonal carcinoma"

/clone\_lib="NIH\_MGC\_92"

/lab\_host="DH10B"

/note="Vector: pCMV-SPORT6"

1..2016

/gene="ERG"

/db\_xref="LocusID:2078"

/db\_xref="MIM:165080"

104..1543

/codon\_start=1

/product="v-ets erythroblastosis virus E26 oncogene like, isoform 1"

/protein\_id="AAH40168.1"

/db\_xref="GI:25304065"

/translation="MASTIKKALSVSDQSLFECAYGTPHLAKTMTASSSDYQQT SKMSPRVQQDWSLQPPARVTIKMCNPSQVNSRNSPDCSVAGKXMGVSPDVGVM NYGSYMEKHMPPNMTNERRVIVADPTLMTSDHVRQWLEWAVKYLDPVNNILF QNIDKELCKMTKDDFQRLTYSYNADILLSHLYLREPLHLTSDVDKALQNSPRL WHARNTGGAARIFPNTSVYPRATORTRPDLVPYPPRRSAWTHGHTPQSKAAQPS PSTVPTKDRQPLDPYAILGPTSSRLANRSGSQIQWFLLELLSDSSNSCIWEG TNGEFMTDTPDEVARWGRGSKPNNTDKLSRALRYYYDKNIMTKVGRKYAKFDF HGIAQALQHPHPSSELYKPSDLPYMGSIHAHPQGNFVAPHPPALPVTSSSFFAAPN PYWNSPTGGIYENTRLPTSHMPSHLGTYT"

446..694

/misc\_feature

/note="SAM\_PNT; Region: Sterile alpha motif (SAM)/Pointed domain"

/db\_xref="CDD:pfam02198"

1031..1282

/misc\_feature

/note="Ets; Region: Ets-domain"

/db\_xref="CDD:pfam00178"

ORIGIN

Alignment Scores:

Pred. No.: 4,75E-153 Length: 2016

Score: 2313.00 Matches: 429

Percent Similarity: 91.44% Conservative: 9

Best Local Similarity: 89.56% Mismatches: 13

Query Match: 94.64% Indels: 28

DB: 9 Gaps: 2

US-09-902-772-2 (1-451) x BC040168 (1-2016)

Qy 1 MetAlaSerThrIleLysGluAlaLeuSerValSerGluAspGlnSerLeuPheGlu 20

Db 104 ATGCCAGCACTATTAAAGGAAGCTTATCAGTTGTAGTGAGGACCAGTCGTGTTGAG 163

Qy 21 CysAlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGlu 40

Db 164 TGTGCTTACGACGACCCACCTGGCTAAGACAGATACCGCGTCTCTCCACCGAC 223

Qy 41 TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnPro 60

Db 224 TATGGACAGACTTCCAAAGATGAGCCACGCGTCCCTCAGCAGGATGGTGTCTCAACC 283

Qy 61 ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80

Db 284 CCAGCCAGGGTCCACCATCAAAATGGAATGTAAACCTAGCCAGGTGAATGCTCAAGAAC 343

Qy 81 SerProAspAspCysSerValAlaLysGlyGlyLysMetValSerSerSerAspAsnVal 100

Db 344 TCTCCTGATGAATGCAGTGTGGCCAAAGCGGGAAGATGTTGGCGGACCCACGACCGTT 403



```

QY 42 GlyGlnThrSerLysMetSerProArgValProGlnAspTrpLeuSerGlnProPro 61
Db 292 GGCAGAGATCAAGATGAGTCCAGAGTCCCTCAGCAGGACTGGCTGTCTCAAGCCCA 351
QY 62 AlaArgValThrLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSer 81
Db 352 GCCAGGGTCCACATCAAGATGGAGTCAACCTTAGTCAGGTGAATGGTTCAGGAATCA 411
QY 82 ProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsnValGly 101
Db 412 CCTGATGAGTGCAGTGTGAACAAAGGTGGGAAGATGGTGGCAGCCGGATACGTGGG 471
QY 102 MetAsnTrpGlySerTrpMetGluGlnLysHisLeProProProAsnMetThrAsn 121
Db 472 ATGACTACGGCAGCTACATGAGGAGAGCATGTGCCGCTCCCAATATGACACAAAT 531
QY 122 GluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGln 141
Db 532 GAGCCAGAGTATCGTCCCTGCGATCTTACTCTGTGGAGCAGACCATGTCGGACAG 591
QY 142 TrpLeuGluTrpAlaValLysGluTrpGlyLeuProAspValAspIleLeuLeuPheGln 161
Db 592 TGGCTGGAGTGGCGGTGAAGAATATGCGCTCTCGATGTGGAGCTTACTATTTCAG 651
QY 162 AsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThrPro 181
Db 652 AATATCGATGGGAAGAGCTGTGCAAGATGACAAAGGATGACTTCCAGCGGCTCACGCG 711
QY 182 SerTrpAsnAlaAspIleLeuLeuSerHisLeuHisTrpLeuArgGluArg----- 198
Db 712 AGCTACATGCCAGATCTTCTCTCACATCTCCACTACCTCAGAGAGACTCCCTTCCA 771
QY 198 ----- 198
Db 772 CATCTGACTTCGATGAGCTTGATAAGCTTTACAAACTCTCCACGGTTAATCATGCC 831
QY 199 -----GlyValaThrPheIlePheProAsnThrSerValTrpGluAlaThr 214
Db 832 AGAAACACAGGGGTGCAGCTTTATTTTCCCAATATCTTCAGTATATCCGAGAGCTACG 891
QY 215 GlnArgIleThrArgProAspLeuProTrpGluGlnAlaArgArgSerAlaTrpThr 234
Db 892 CAAGAATATCAACTAGGCCAGATTACCTTATGAGCCTCCAGGAGATCAGCGCTGGACC 951
QY 235 SerHisSerHisProThr---GlnSerLysAlaThrGlnProSerSerSerThrValPro 253
Db 952 GGCACAGCCCTCACCCTCAGTCCAAAGCTGCTCAGCCATCTCCCTCTGCAAGTGGCC 1011
QY 254 LysThrGluAspGlnArgProGlnLeuAspProTrpGlnIleLeuGlyProThrSerSer 273
Db 1012 AAACTGAAGACCGCGTCTCAGTTAGATCTTACCAGATCTCTGGGACCGACCATGAGC 1071
QY 274 ArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeu 293
Db 1072 CCCTTGCTTAATCCAGGTAGTGGCAGATCCAGCTGTGGCAGTCTCTCTCGAACTCCTG 1131
QY 294 SerAspSerSerAspAsnCysIleThrTrpGluGlyThrAsnGlyLysPheLysMet 313
Db 1132 TCAGACAGCTCCAACTCCACTGATCATCTCTGGGAGGACCAACAGCGGAGTTCAGATG 1191
QY 314 ThrAspProAspGluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMetAsn 333
Db 1192 ACAGACCCGAGCAGGTGGCTCGCGCTCGGGGAGAGAGAGAGCAAGCCCAACATGAAC 1251
QY 334 TyrAspLysLeuSerArgAlaLeuArgTrpTrpTrpAspLysAsnIleMetThrLysVal 353
Db 1252 TATGACAGCTCAGCGCGCCCTCCGCTACTACTACGACAAACATCATGACCCAGGTG 1311
QY 354 HisGlyLysArgTrpAlaTrpLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnPro 373
Db 1312 CACGGGAAGCGCTACGCTTACAGTTGACTTCCACGGGATTCGCCAGCGCTGCAGGCC 1371

```

```

QY 374 HisProProGluSerSerMetTyrLysTrpProSerAspLeuProTrpMetSerSerTrp 393
Db 1372 CACCCTCTCAGTGCCTGCTACAGTACCGCTCCGACTGCATACATACATGCGCTCTAT 1431
QY 394 HisAlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProValThr 413
Db 1432 CAGCCACCCAGAGATGACTTTGTCTCCACCCTCCCGCTCTCCAGTACACA 1491
QY 414 SerSerSerPhePheAlaAlaProAsnProTrpTrpAsnSerProThrGlyGlyLeuTrp 433
Db 1492 TCTTCCAGTTCTTCTGCTTCCCGAACCATGCTGCAATTCACCGCTGGGCGCATCTAC 1551
QY 434 ProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTrpTrp 451
Db 1552 CGAACACTAGGTCCTCCAGCCAGCATATGCGCTCTCACCCTGGGACCTACTAC 1605

RESULT 8
LOCUS XLAJ4126 2012 bp mRNA linear VRT 21-JAN-2000
DEFINITION Xenopus laevis erg gene (erg_E).
ACCESSION AJ224126
VERSION AJ224126.1 GI:5420047
KEYWORDS ERG gene; transcription factor.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1
AUTHORS Baltzinger,M., Mager-Heckel,A.M. and Remy,P.
TITLE X1 erg: expression pattern and overexpression during development
JOURNAL Dev. Dyn. 216 (4-5), 420-433 (1999)
MEDLINE 20093678
PUBMED 10633861
REFERENCE 2 (bases 1 to 2012)
AUTHORS Baltzinger,M.
TITLE Direct Submission
JOURNAL Submitted (24-FEB-1998) Baltzinger M., UPR9005, MPMCD, Centre
National de la Recherche Scientifique, 15, RUE Rene Descartes,
67084, FRANCE
FEATURES
source
1..2012
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="lambda ZAPII-erg E"
/clone_lib="lambda ZAPII"
/dev_stage="stage 30 embryo"
1..2012
/gene="erg"
25..1595
/gene="erg"
/function="transcription factor"
/codon_start=1
/protein_id="CAB46567.1"
/db_xref="GI:5420048"
/db_xref="GOA:Q9W629"
/db_xref="SPTREMBL:Q9W629"
/translation="MIOCTVDPSSAHIKKALSVVSDQSLFECTYGTGPHLTKTMTASS
SDYGTGSKMSPRVPODNLSDPPSVTIKMCNPSOVNGSRSSPDGCSIGKSGKMG
GSDVGMVGSVTEEXHISPPNMTNERSVIYPADPTLWSTDHVRQWLKWAKEYGLP
DVDFLQNDIGKELCKMTKEDFHLTPSYNADILLSHLYLEERGASRIFPNSVYQ
DANQRIPSKQDLSYSPRSANWHPPSKASQPSSTTVFKTEDPRQDLPQILGPT
SERLANPGSQIQLWQFLLELLSDSNCSITWEGTNGBFKMDPDEVARRWGRSK
PNMYDKLGRALRYIKYDKNIMTKHGKRYAKFDHGIQAQALQPHPESTMYKPSBL
PYMSVYHAHPQKNFVAPHPALPVTSSFFFAAPNAYMNSPTGSIYPTNRLPASHMS
HLGTYY"
ORIGIN
Alignment Scores: 1.46e-144 Length: 2012
Pred. No.: 2192.00 Matches: 402
Score:

```

```

Percent Similarity: 94.63%      Conservative: 21
Best Local Similarity: 89.93%    Mismatches: 22
Query Match: 89.69%             Indels: 2
DB: 5                           Gaps: 2

US-09-902-772-2 (1-451) x XLAJ4126 (1-2012)

QY 5 IleLysGluAlaLeuSerValIleSerGluAspGlnSerLeuPheGluCysAlaTyrGly 24
DB 258 ATCAAGAAGCTCTTTTCAGTGGTGGTGAAGATCAGTCTGTTATTTCGAGTGCACTACGGA 317
QY 25 SerProHisLeuAlaLysThrGluMetThrAlaSerSerSerSerSerGluTyrGlyGlnThr 44
DB 318 AGCCACATCTTACTAAACGGAGATGACCGCATCTTCTCCAGTACTATGGCAAAACC 377
QY 45 SerLysMetSerProArgValProGlnGlnAspThrLeuSerGlnProProAlaArgVal 64
DB 378 TCGAAATGAGTCCACGCTGCTGCTCAACAGACTGGCTCTCACAGCCTCCATCCAGGGTG 437
QY 65 ThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSerProAspAsp 84
DB 438 ACATCAAGATGGATGCAACCTAGTCAAGTCAATGATCAAGGAGCTCCCGGATGAC 497
QY 85 CysSerValAlaLysGlyLysMetValSerSerSerSerSerSerSerValGlyMetAsnTyr 104
DB 498 TGCAGTATAGGAAAGGAAGCAAAATGGGTGGAGGTCGGATAATGTTGGAATGAATAT 557
QY 105 GlySerTyrMetGluGluLysIleProProProAsnMetThrThrAsnGluArgArg 124
DB 558 GGCAGCTACATAGAGAAACACATTTCTCCCCCAACATGACAAACATGAGAGAGA 617
QY 125 ValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGlnTrpLeuGlu 144
DB 618 GTGATTGACCTGCTGACCTACTTGTGGAGCACAGATCATGTCCAGACAGTGGCTAGAG 677
QY 145 TrpAlaValLysGluTyrGlyLeuProAspValAspIleLeuLysPheGlnAsnIleAsp 164
DB 678 TGGGCATTAAGATACCGCTTCCAGATGTCGATGTGCTCTGTTCCAAACATCGAC 737
QY 165 GlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThrProSerTyrAsn 184
DB 738 GCGAGGATATGCAAGATGACCAAGAGAGATTTCCATCGACTCACGCCAAGCTTACAT 797
QY 185 AlaAspIleLeuSerHisLeuHisTyrLeuArgGluArgGlyAlaThrPheIlePhe 204
DB 798 GCTGATATCTCTCTCGCATCTTCACTACCTCAGAGAAAGGTGCTCTTTTATTATTC 857
QY 205 ProAsnThrSerValTyrProGluAlaThrGlnArgIleThrThrArgProAspLeuPro 224
DB 858 CCAACTCATCGTTTATCAAGCGCAACCAAGANATCCCGAGCAGCAGATTTATCA 917
QY 225 TyrGluGlnAlaArgSerAlaTrpThrSerHisSerHisProThrGlnSerLysAla 244
DB 918 TATGAGCCCTCAAGGAGATCTGTTGGACAAATCATCTCTGCACCA---CCATCAAAAGCA 974
QY 245 ThrGlnProSerSerSerThrValProLysThrGluAspGlnArgProGlnLeuAspPro 264
DB 975 TCTCAACCA---TCTACCAAGATTTCCAAACAGAGAGCCCGAGCCACAGCTAGATCT 1031
QY 265 TyrGlnIleLeuGlyProThrSerArgLeuAlaAsnProGlySerGlyGlnIleGln 284
DB 1032 TATCAGATCTTGGGCAACAGCAGCCGACTTGCAAATCCAGGAAGTGGCGAGATTCCAG 1091
QY 285 LeuTrpGlnPheLeuLeuGluLeuSerAspSerSerAsnSerAsnSerAsnSerAsnSer 304
DB 1092 CTTCTGCAATCTTCTAGCACTGCTTTCCGATAGTTCCTCACTCACTGCACTTACTTGG 1151
QY 305 GluGlyThrAsnGlyGluPheLysMetThrAspProAspGluValAlaArgArgTrpGly 324
DB 1152 GAAGGAACCAATGGAGATTTAAGATGACCGGATCTCTGATGAAGTGGCAGAGCTTGGGG 1211
QY 325 GluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArgAlaLeuArgTyrTyr 344

```

---

```

1212 GAGAGGAAAGCAACCAACCAACATGATATGATAAGCTCAGCGTGCACCTCGTTACTTAC 1271
345 TyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaTyrLysPheAspPhe 364
1272 TAGCATAAATAATATTATGACTAAAGTCCATGCAAGCGCTATGCTTACAAATTTGATTTT 1331
365 HisGlyIleAlaGlnAlaLeuGlnProHisProGluSerSerMetTyrLysTyrPro 384
1332 CATGGATTTGCTCAGCTCTTCAACTCATCTCCAGATCAACTATGTACAGTATCC 1391
385 SerAspLeuProTyrMetSerSerTyrHisAlaHisProGlnLysMetAsnPheValAla 404
1392 TCAGAACTTCCATATATGAGCTCATACCATGACATCCACAGAGATGAACCTTTGTGTC 1451
405 ProHisProProAlaLeuProValThrSerSerSerPhePheAlaAlaProAsnProTyr 424
1452 CCCATCCCGCAGCTTACCTGTGACATGCTCAAGTTCTTTGCGACCCCTAATGCATAC 1511
425 TrpAsnSerProThrGlyGlyIleTyrProAsnThrArgLeuProAlaAlaHisMetPro 444
1512 TGGAAATTCACCAACTGGAAGTATTTATCCAAATACTCGGCTGCCAGCTAGCCATATGCT 1571
445 SerHisLeuGlyThrTyrTyr 451
1572 TCTCATCTTGGAACTTACTAT 1592

RESULT 9
HUMERG2
LOCUS HUMERG2 3166 bp ss-RNA linear PRI 08-NOV-1994
DEFINITION Human erg2 gene encoding erg2 protein, complete cds.
ACCESSION MI7254.1
VERSION MI7254.1 GI:182186
KEYWORDS erg 2 protein.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3166)
AUTHORS Rao, V.N., Papas, T.S. and Reddy, B.S.
TITLE erg, a human ets-related gene on chromosome 21: alternative
splicing, polyadenylation, and translation
JOURNAL Science 237 (4815), 635-639 (1987)
MEDLINE 87263429
PUBMED 3299708
COMMENT Original source text: Human cell line COLO 320, cDNA to mRNA, clone
lambda 12.
FEATURES
source 1..3166
/organism="Homo sapiens"
/mol_type="genomic RNA"
/db_xref="taxon:9606"
/map="21q22.3"
257..1645
/genes="ERG"
257..1645
/genes="ERG"
/notes="erg 2 protein"
/codon_start=1
/protein_id="AA52398.1"
/db_xref="GI:182187"
/db_xref="GDB:G00-119-884"
/translation="MTQVDPAAHFKELSVSVSEDSQSLFECAIYGPPLAKTENTASS
SDYGGTSGMSPRVPOQWLSQSPARVTIMECNPSQVNSRNSPDECSVAKGKRVG
SPVTGNTGSGYMEKEKMPFNNMTTNERAVIYPADPTLWSTDRVQWLEWAVEYGLP
DVNILLFQNTDGKELCKMDQFRLTPSYNADILLSHLYLFEETPLHLTSDVDKA
LQNSPLRMAHNTDLPVEPPRSATWGHTGHPQSKAAQSPSPSTVPKTEQDRLQDPY
OILGPTSSRLANPGSGIOLWFLLELLSDSSNSCTIWEGTGFKMTDPDEVARRW
GERKSPNNYDKLSALARYIDKNTKTVHGKRYAKPDFHGIQALQPHPESSLY
KYPSDLPYMGSIYHAHFQKQNFVAPHPFALPVTSSTSPFAAPNPNYWSFTGLIYNTRLP
TSHMPGHLGTY"
repeat_region 291..298
/notes="8-bp repeat sequence"
repeat_region 509..516

```



polya\_signal /note="g-bp repeat sequence" 3143..3149  
 ORIGIN 227 bp upstream of PvuII site; chromosome 21.  
 Alignment Scores:  
 Pred. No.: 1.25e-143 Length: 3166  
 Score: 2182.00 Matches: 407  
 Percent Similarity: 91.72% Conservative: 14  
 Best Local Similarity: 88.67% Mismatches: 24  
 Query Match: 89.28% Indels: 14  
 DB: 9 Gaps: 4  
 US-09-902-772-2 (1-451) x HUMERG2 (1-3166)  
 QY 2 AlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGluCys 21  
 DB 281 CGAGTCTATACAGAGACCTTATCAGTTGTGAGTGGACCAAGTCTGTTTGGATGT 340  
 QY 22 AlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGluTyr 41  
 DB 341 GCCTACGGAAGCCACACCTGGCTAAGACAGAGATGACCGGTCTCTCTCCAGGACTAT 400  
 QY 42 GlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnProPro 61  
 DB 401 GGACAGACTTCCAGATGAGCCCGCCGCTCCCTCAGCAGGATGGCTGTCTCAACCCCCA 460  
 QY 62 AlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSer 81  
 DB 461 GCCAGGTCACCATCAAAATGGATGTAAACCTAGCCAGGTGAATGGCTCAAGGAATCT 520  
 QY 82 ProAspAspCysSerValAlaLysGlyGlyLysMetValSerSerSerAspAsnValGly 101  
 DB 521 CCTGATGAATGCAGTGTGGCCAAAGCGCGGAAGATGGTGGCGACGCCACACCGTTGGG 580  
 QY 102 MetAsnTyrGlySerTyrMetGluGluLysHisIleProProProAsnMetThrThrAsn 121  
 DB 581 ATGAATCTACGGAGCTCATGGAGGAGAGAGACATGCGCCACCCCAACATGACCGAGAC 640  
 QY 122 GluArgArgValIleValProAlaAspProThrLeuTyrSerThrAspHisValArgGln 141  
 DB 641 GAGCGCAGAGTTATCTGTCGACGACATCTACGCTATGGAGTACAGACCATGTGCGGCG 700  
 QY 142 TrpLeuGluTyrAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPheGln 161  
 DB 701 TGGCTGGAGTGGCGGTGAAGANATATGGCTTCCAGAGCTCAACATCTGTTATTCCAG 760  
 QY 162 AsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThrPro 181  
 DB 761 AACATCGATGGAGAGAACTGTGCAAGATGACCAAGGACGACTTCCAGAGGCTCACCC 820  
 QY 182 SerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArgGlyAlaThr 201  
 DB 821 AGCTACACGCGGACATCTCTCTCATCATCTCCACTACCTCAGAGAGACTCCCTCTCCA 880  
 QY 202 PheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrThrArgPro 221  
 DB 881 CATTG-----ACITCA-----GATGATGTTGATAAGCGCTTACAAACTCTCCA 925  
 QY 222 -----AspLeuProTyrGluGlnAlaArgArgSerAlaTyr 233  
 DB 926 CGGTTAATGTCATGATAGAACACAGATTATACCATATGAGCCCGCCAGGAGATCAGCCTG 985  
 QY 234 ThrSerHisSerHisProThr--GlnSerLysAlaThrGlnProSerSerSerThrVal 252  
 DB 986 ACCGTACCGCCACCCACCGCCGAGTGCAGAAAGTGTCAACCATCTCTCTTCCACAGTG 1045  
 QY 253 ProLysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSer 272  
 DB 1046 CCCAAACTGAAGACGAGCGTCTCAGTAGATCTCTATCATGATTTCTGGACCAACAGT 1105  
 QY 273 SerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTyrGlnPheLeuLeuGluLeu 292

Db 1106 AGCCGCTTGCATAATCCAGGCAGTGGCCAGATCCAGCTTTGGCAGTTCCTCTGGAGCTC 1165  
 QY 293 LeuSerAspSerSerAsnSerAsnSerCysIleThrTyrGluGlyThrAsnGlyGluPheLys 312  
 Db 1166 CTGTGGGACAGTCTCCAACTCCAGCTCATCATCTGGGAAGGACCAACGGGGAGTTCAAG 1225  
 QY 313 MetThrAspProAspGluValAlaArgAgtTyrGlyGluArgLysSerLysProAsnMet 332  
 Db 1226 ATGACGGATCCCGACGAGGTGGCCGGCGCTGGGGAGAGCGGAAGCAAAACCAACATG 1285  
 QY 333 AsnTyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLys 352  
 Db 1286 AACTACGATAAGCTCAGCCCGCCCTCCGTACTACTATGACAAGAACATCATGACCAAG 1345  
 QY 353 ValHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGln 372  
 Db 1346 GTCCATGGGAGCGCTACGCTACAGTTCCGATTCACCGGATCGCCAGGCGCTCCAG 1405  
 QY 373 ProHisProProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSer 392  
 Db 1406 CCCACCCCGGAGTCACTCTGTACAAAGTACCCCTCAGACCTCCCGTACATGGGCTCC 1465  
 QY 393 TyrHisAlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProVal 412  
 Db 1466 TATCAGCCCGCCACAGAGATGAATTTGTGGCGCCCGCCACCTCCAGCCCTCCCGTG 1525  
 QY 413 ThrSerSerSerPhePheAlaProAsnProTyrTyrAsnSerProThrGlyGlyIle 432  
 Db 1526 ACATCTCCAGTTTGTGTCGCCCAACCCCATCTGGAATTCACCAACTGGGGGTATA 1585  
 QY 433 TyrProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451  
 Db 1586 TACCCCAACACTAGGCTCCCGACCGCAATAGCTTCTCATCTGGGCACTTACTACTAC 1642  
 RESULT 10  
 LOCUS AB031088 1808 bp mRNA linear ROD 07-AUG-2001  
 DEFINITION Rattus norvegicus VESP14 mRNA for vascular endothelial cell specific protein 14, complete cds.  
 ACCESSION AB031088.1 GI:15128488  
 VERSION vascular endothelial cell specific protein 14.  
 KEYWORDS Rattus norvegicus (Norway rat)  
 SOURCE Rattus norvegicus  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 REFERENCE 1 (bases 1 to 1808)  
 AUTHORS Aoki,T., Toyoda,H., Nishimoto,S., Tawara,J., Ukai,Y. and Komurasaki,T.  
 TITLE Identification of VESP14,a vascular endothelial cell specific protein  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 1808)  
 AUTHORS Aoki,T., Toyoda,H., Nishimoto,S., Tawara,J., Ukai,Y. and Komurasaki,T.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-AUG-1999) Tomomi Aoki, Taisho Pharmaceutical Co., Ltd., Medicinal Research Laboratories, Molecular Biology Lab., 1-403, Yoshino-cho, Omiya, Saitama 3308530, Japan (E-mail:sl7704@ccm.taisho.co.jp, Tel:048-663-1111, Fax:048-652-7254)  
 FEATURES  
 source 1..1808  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10116"  
 /tissue\_type="liver"  
 gene 1..1808  
 /genes="VESP14"  
 CDS 63..1430  
 /genes="VESP14"  
 /note="similar to human erg1"

```

/codon_start=1
/product="vascular endothelial cell specific protein 14"
/protein_id="BAB62744.1"
/db_xref="GI:15128489"
/translaton="MASTIKKALSVSKQDLSFECAYGTPLHAKTEMTASSSDYDGT
SKMSPRVQDQWLSQPARVTIKMECPNSOVNGSRNSPDECSVTGKGMVGSPTDVTGM
SYGSMKEKMPNPNTNRRVIVPADPTLMSDTHVROWLEWAVEYGLDLDVLLF
QNIQKELCKMTKDDFORLTPSYNADILLSHLYLRETPLPLHTSDVDKALQNSRL
MRANITDLPFPPRSWTGSHSTFQSKAAQSPSTPKTEDEQFQDLPYQILGPTS
SHLANPQSGQQLWQFLELLSDSSNSNCITWETNGTFRONTDFDEVARWGRKSKP
NMNYDKLSRALRYFDKNIWKVKRYAYKDFPHGAQALPHPESSLYKYPSPDL
YMGSHYTHPQKMFVAPHPALPVTSSFFATNPYMNPTGGIYENTRLPASHMPSH
LGTYY"

```

## ORIGIN

## Alignment Scores:

```

Pred. No.: 1,05e-143 Length: 1808
Score: 2179.00 Matches: 406
Percent Similarity: 91.30% Conservative: 14
Best Local Similarity: 88.26% Mismatches: 26
Query Match: 89.16% Indels: 14
DB: 10 Gaps: 4

```

US-09-902-772-2 (1-451) x AB031088 (1-1808)

```

QY 1 MetAlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20
Db 63 ATGCCAGCAGCTATTAAAGAGCGCTTGTCAGTTGTGAGCAAGGACCACTACTATTGAG 122
QY 21 CysAlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGlu 40
Db 123 TGTGCTACGGAACCCACACCTGCTAAGACAGAAATGACCGCATCTCTTCAGGTGAC 182
QY 41 TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnPro 60
Db 193 TATGCCAGACATCCAGATGATGCTCCAGATGCCACACAGGACTGGCTGTCTCAGCCC 242
QY 61 ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
Db 243 CCAGCCAGGCTCACCATCAAGATGAGTGCACCACTAGCCAGGTGAACGGTTCAGGAAC 302
QY 81 SerProAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsnVal 100
Db 303 TCACCTGACGAATGATGATGACCAAGTGGAGATGGTGGCAGCGCTGATCTGTG 362
QY 101 GlyMetAsnTyrGlySerTyrMetGluGlyLysHisIleProProProAsnMetThr 120
Db 363 GGAATGAGCTACGGCAGCTACATGAGGAGAGACACATGCGCGCCCAATATGACCACG 422
QY 121 AsnGluArgArgValIleValProAlaAspProThrLeuTyrSerThrAspHisValArg 140
Db 423 AATGAACGAGAGTATGCTGCTCCAGATCTCTCTGTGAGCAGACCAATGTCGG 482
QY 141 GlnTyrLeuGluTyrAlaValLysGluTyrGlyLeuProAspValAspIleLeuPhe 160
Db 483 CAGTGGCTGAGTGGGAGTGAAGAAATAGCGCTCTAGACGTGGAGCTCTATTATT 542
QY 161 GlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThr 180
Db 543 CAGAATATTGACGGGAGAGCTGTGCAAGATGACAAAAGATGACTTCCAGAGGCTCACT 602
QY 181 ProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArgGlyAla 200
Db 603 CCGAGCTACAAAGCCGACATCTTCTCTCAGATCTCCACTACTCTCAGAGACTCCCTT 662
QY 201 ThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrArg 220
Db 663 CCACATCTG-----ACATCC-----GACGACGTTGTAAGCGCTTTCAAAACTCT 707
QY 221 Pro-----AspLeuProTyrGluGlnAlaArgSerAla 232
Db 708 CCAGGTTAATGATCTAGAAACACAGATTTACTTATGAGCTCCCGGAGATCAACC 767

```

```

QY 233 TrpThrSerHisSerHisProThr---GlnSerLysAlaThrGlnProSerSerSerThr 251
Db 768 TGGACGGGCGCACAGCCACCCCTCAGTCCAAAGCTGCCAGCAATCTCCCTCCACA 827
QY 252 ValProLysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThr 271
Db 828 GTGCCCAAACTGAAGACCAGCGTCCTCAGTTAGATCTTACCAATCTCTGGGACCCACC 887
QY 272 SerSerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTyrGlnPheLeuGlu 291
Db 888 AGTAGTCGCTTGTCTAAATCCAGTAGTGCACAGATCCAGCTGTGCAGTCTCTGCTAGNA 947
QY 292 LeuLeuSerAspSerSerAsnSerAsnCysIleThrTyrGluGlyThrAsnGlyGluPhe 311
Db 948 CTCCTGTCTGACAGCTCCAACTCCAACTGCATCACTGGGAAGGACCAACAGGGAGTTC 1007
QY 312 LysMetThrAspProAspGluValAlaArgTyrGlyGluArgLysSerLysProAsn 331
Db 1008 AAGATGACAGACCCGAGTAGGTGCTCGGCTCGGGGAGCGGAGAGGAGCCCAAC 1067
QY 332 MetAsnTyrAspLysLeuSerArgAlaLeuArgTyrTyrAspLysAsnIleMetThr 351
Db 1068 ATGAACTATGACAACTCAGCGCTGCTCCCTCTACTACTACGAGAAAACATCATGACC 1127
QY 352 LysValHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeu 371
Db 1128 AAGGTGCACGGGAAGCGCTATGCTTACAAAGTTTGACTTCCACGGGATGCCAGGCCCTG 1187
QY 372 GlnProHisProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSer 391
Db 1188 CAGCCCACTCCCGGAGTCTGCTGTACAGTACCTCCCTCCGACCTGCGGTACATGGGC 1247
QY 392 SerTyrHisAlaHisProGlnLysMetAsnPheValAlaProHisProAlaLeuPro 411
Db 1248 TCTATCACACACCCCGCAGAGATGAATTTGGCTCCCGCTCCCGCTCCCGCTCCCA 1307
QY 412 ValThrSerSerSerPhePheAlaProAsnProTyrTyrAsnSerProThrGlyGly 431
Db 1308 GTCAATCTTTCAGTTTCTTGTCTACCCCGAACCATCTAGAAATTCGCGGCTGGGGC 1367
QY 432 IleTyrProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
Db 1368 ATCTACCCGAACAGTAGGCTCCGACGACCATATGCTCTCTCAGCTGGGACCTACTAC 1427

RESULT 11
AB073079 2133 bp mRNA linear ROD 17-OCT-2001
LOCUS Mus musculus Erg mRNA, mouse homolog of Human ets-related gene ERG,
DEFINITION transcript variant 2, complete cds.
ACCESSION AB073079 GI:16197542
VERSION
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 Ozawa,R., Noguchi,H., Taylor,T.D., Takeda,T., Hattori,M. and
Sakaki,Y.
Mus musculus Erg mRNA
Published Only in Database (2001)
REFERENCE
2 (bases 1 to 2133)
Ozawa,R., Noguchi,H., Taylor,T.D., Takeda,T., Hattori,M. and
Sakaki,Y.
Direct Submission
Submitted (16-OCT-2001) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
FEATURES
1. 2133
/organism="Mus musculus"
source

```

```

/mol_type="mRNA"
/db_xref="taxon:10090"
/chromosome="16"
/clone="Erg-2"
1..2133
/gene="Erg"
149..1540
/gene="Erg"
/notes="homolog of human ets-related gene (ERG), transcript
variant 2"
/codon_start=1
/protein_id="BA869949.1"
/db_xref="GI:16197543"
/translation="MIQTVDPAAHKEALSVSVSEDSLEFCAYGTPHIAKTEMTASS
SSDYQTSKMSRPVQDDWSQAPARTVITKECPQSVGRNSRSPDECNVKGGKMGV
SPDVTGVNGSYMEKEKVPVMTNERRVIVPADPTLWSTHVRQWLEWAVEYGLL
DVLVLLFQNDIGKELCKMTKDDFORLTPSYNADILLSHLHYLRETPHLTSDVDVKA
LQNSPRHARNTGGAFFIPNTSVYPEATORITTPAAQSPSAVPKETEDQPDQDP
YQILGPTSRNLNPGSQIQLWQPLELLSDSNINCITWEGTFGEFMTDPDEVARL
WGERKSPNNWYDKSLRALRYDYDKNIWTKVHKRYAYKDFPHGIAQALQPHPPESL
YKPSDLFVYGSYHAHPQKMFVSPHPFALPVTTSSFPSPFNPNWNSPTGGIYPNTRL
PASHMPSHLGTY"

```

## ORIGIN

```

Alignment Scores:
Pred. No.: 2,97e-142 Length: 2133
Score: 2159.50 Matches: 403
Percent Similarity: 87.00% Conservative: 12
Best Local Similarity: 84.49% Mismatches: 13
Query Match: 88.36% Indels: 49
DB: 10 Gaps: 2

```

US-09-902-772-2 (1-451) x AB073079 (1-2133)

```

QY 2 AlaserThrileLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGluCys 21
DB 173 GCAGCTCATATTAGGAGGCGCTTGTCAGTTGTGAGCGAGGACCAGTCATTATTGAGTGT 232
QY 22 AlaTyGlySerProHisLeuAlaIysThrGluMetThrAlaSerSerSerSerGluTy 41
DB 233 GCCTACGGAAGCGCCACACTGCTGAAGACAGAGATGACCGCATCTCTTCCAGTGACTAT 292
QY 42 GlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnProPro 61
DB 293 GGCCAGACATCAAGATGAGTCCAGAGTCCCTCAGCAGGAGTGGCTGTCTCAAGGCCCA 352
QY 62 AlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSer 81
DB 353 GCCAGGGTCACCATCAAGATGAGTGCAACACCTAGTCAGGTGAATGGTTCCAGGAATCA 412
QY 82 ProAspAspCysSerValAlaLysGlyGlyLysMetValSerSerSerSerAspAsnValGly 101
DB 413 CTTGATGAGTGAGTGAGTGAACAAAGTGGAGAGATGGTGGCAGCCCGGATACGTGGGG 472
QY 102 MetAsnTyGlySerTyMetGluGluLysHisIleProProProAsnMetThrThrAsn 121
DB 473 ATGAGTACGGCAGCTACATGAGGAGGAGGAGATGTCGCGCTCCCAATATGACCAACAAT 532
QY 122 GluArgGluValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGln 141
DB 533 GAGCGCAGAGTATCGTCCCTCAGATCCTCTCTGTGAGCAGCAGACATGTCGACAG 592
QY 142 TrpLeuGluTrpAlaValLysGluTyGlyLeuProAspValAspIleLeuLeuPheGln 161
DB 593 TGGCTGGAGTGGCGGTGAAGAATATGCGCTCCCTCGATGGACGTCTTACTATTTCAG 652
QY 162 AnnilaAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThrPro 181
DB 653 AATATCGATGGGAAGAGCTGTGCAAGATGACAAAGGATGACTCCAGCGGCTCAGCCG 712
QY 182 SerTyAsnAlaAspIleLeuLeuSerHisLeuHisTyLeuArgGluArg 198
DB 713 AGCTACAAATGCGGACATCTTCTTCACATCTCCACTACCTCAGAGAGACTCCCTTCCA 772

```

```

QY 198 ----- 198
DB 773 CATCTGACTTCGATGACGTTGATAAGGCTTTACAAAACCTCTCCACGGTTAATGATGCC 832
QY 199 -----GlyValaThrPhePheProAsnThrSerValTyProGluAlaThr 214
DB 833 AGAAACACAGGGGGTGCAGCTTTTATTTTCCCAATATCTTCAATATATCCGAGCTAGC 892
QY 215 GlnArgIleThrThrArgProAspLeuProTyGluGlnAlaArgSerAlaTrpThr 234
DB 893 CAAAGAAATTACAACTAGGCCA----- 913
QY 235 SerHisSerHisProThrThrGlnSerLysAlaThrGlnProSerSerSerThrValProLys 254
DB 914 -----GCTGCTCAGCCATCTCCCTCTGCAGTGCCTCAAA 946
QY 255 ThrGluAspGlnArgProGlnLeuAspProTyGlnIleLeuGlyProThrSerSerArg 274
DB 947 ACTGAAGACCAGGCTCTCAGTTAGATCTTACCAGATCTCTGGGACCGACCACTAGTACCGC 1006
QY 275 LeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuLeuLeuSer 294
DB 1007 CTTGTCTAATCCAGGTAGTGGCCAGATCCAGCTGTGGCAGTTCTCTGCTCGAATCTCTGTCA 1066
QY 295 AspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThr 314
DB 1067 GACAGCTCCAACTCCAACTGCACTCCTGGGAGGACCAACACGGGGAGTTCAAGATGACA 1126
QY 315 AspProAspGluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMetAsnTy 334
DB 1127 GACCCGACGAGGTGGCTGGGGGAGAGAGAGCAAGCCCAACATGAACATAT 1186
QY 335 AspLysLeuSerArgAlaLeuArgTyTrpTyAspLysAsnIleMetThrLysValHis 354
DB 1187 GACAGCTCAGCGCGCCCTCGCTACTACTACGACAAACATCATGACCAAGGTGCAC 1246
QY 355 GlyLysArgTyAlaTyLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHis 374
DB 1247 GCGAAGCGGTACGCTTACCAAGTTTGACTTCCACGGGATTGCCAGGCCCTGCGAGCCCCAC 1306
QY 375 ProProGluSerSerMetTyLysTyTrpProSerAspLeuProTyTrpMetSerSerTyHis 394
DB 1307 CTTCTCTGAGTCGCTCTGACAGATACCCCTCGACCTGCCATACATGGGCTCTATCAC 1366
QY 395 AlaHisProGlnLysMetAsnPheValAlaProHisProAlaLeuProValThrSer 414
DB 1367 GCCCACCCCGAGAGATGAACCTTGTGTCTCCGCCACCTCTCCGCTCTCCAGTCACTCT 1426
QY 415 SerSerPhePheAlaAlaProAsnProTyTrpAsnSerProThrGlyGlyIleTyPro 434
DB 1427 TCCAGTTTCTTGTCTTCCCGAACCCATCTGGAATTCACGACTGGGGGCATCTACCGC 1486
QY 435 AsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyTy 451
DB 1487 AACACTAGGCTCCAGCCAGCCATATGCTCTCTCACCTGGGCACCTACTAC 1537
RESULT 12
XLAJ4125 1798 bp mRNA linear VRT 21-JAN-2000
LOCUS Xenopus laevis erg gene (erg_A).
DEFINITION AJ224125
ACCESSION AJ224125.1 GI:5420045
VERSION ERG gene; transcription factor.
KEYWORDS Xenopus laevis (African clawed frog)
SOURCE Xenopus laevis
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1
AUTHORS Baltzinger, M., Mager-Heckel, A.M. and Remy, P.
TITLE xl erg: expression pattern and overexpression during development
plead for a role in endothelial cell differentiation

```

JOURNAL Dev. Dyn. 216 (4-5), 420-433 (1999)  
 MEDLINE 20099678  
 PUBMED 10633861  
 REFERENCE 2 (bases 1 to 1798)  
 AUTHORS Balzinger, M.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-FEB-1998) Balzinger M., UPR9005, MDCD, Centre  
 National de la Recherche Scientifique, 15, RUE Rene Descartes,  
 67084, FRANCE

FEATURES Location/Qualifiers  
 source 1..1798  
 /organism="Xenopus laevis"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:8355"  
 /clone="lambda ZAPII-erg A"  
 /clone\_lib="lambda ZAPII"  
 /dev\_stage="stage 30 embryo"  
 1..1798  
 /gene="erg"  
 321..1778  
 /gene="erg"  
 /function="transcription factor"  
 /codon\_start=1  
 /protein\_id="CAB46566.1"  
 /db\_xref="GI:5420046"  
 /db\_xref="GOA:Q9W700"  
 /db\_xref="SPRAMEL:Q9W700"  
 /translation="MQMAFIPTIEDYSNEALSUVSEDSQSLFECTYGTPLHTKTEMTA  
 SSSDYQGMKSPRVQDWSLPSPRTVIRKCNPSQVNGSRSSPDSCIGKSKM  
 GSGSDNVMGYVIEKHSIPNMTNERRVTPADPTLWSTDHVRLQWLWAIKEYG  
 LPDVLFFONIDGKELCKMKEDFHLTPSYNADILLGHLVLRPRLPLHTSDDD  
 KALQNSPLMARHTGASFI PNSSVYODANORIPSDOLSVEPERRSAWTHNPAPP  
 SKASQSTVTPKTEDEPDLDPYCLIGPSSRLANPGSGOIQKQFELLKLLSDSNSN  
 CIWEGNGEFKMTDPEVARNGERKSKPNMYDKLSRALRYNDKNTWTKVGRY  
 AYKDPHGIAQALPQPESTMYKIPSELPYMSYHAHPQGMFVAPHPFPALPVTSSS  
 FFAAPNAYWNSPTGSIYPTRLPASHMSHLGTY"

ORIGIN

Alignment Scores:  
 Pred. No.: 6, 48-142 Length: 1798  
 Score: 2153.50 Matches: 399  
 Percent Similarity: 88.98% Conservative: 21  
 Best Local Similarity: 84.53% Mismatches: 23  
 Query Match: 88.11% Indels: 29  
 DB: 5 Gaps: 3

US-09-902-772-2 (1-451) x XLAJ4125 (1-1798)

Qy 7 GluAlaLeuSerValSerGluAspGlnSerLeuPheGluCysAlaTyrGlySerPro 26  
 Db 366 GAAGCTCTTTTCAGTGTGAGTGAAGATCAGTCGTATTTCAGTGCACCTACGGAACGCCA 425  
 Qy 27 HisLeuAlaLysThrGluMetThrAlaSerSerSerGluTyrGlyGlnThrSerLys 46  
 Db 426 CATCTTACTAAACGGAGATGACCGCATCTTCTCCAGTGCATATGGGCAAACTCGAAA 485  
 Qy 47 MetSerProArgValProGlnGlnAspTrpLeuSerGlnProAlaArgValThrIle 66  
 Db 486 ATGAGTCACCTGTGCTCAACAAGACTGGCTCTCACAGCTCCATCCAGGGTGACCATC 545  
 Qy 67 LysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSerProAspAspCysSer 86  
 Db 546 AAGATGGAATCAACCTCCATCAAGTCAAGTCAAGGAGCTCCCGGATGACTGCAGT 605  
 Qy 87 ValAlaLysGlyLysMetValSerSerSerAspAsnValGlyMetAsnTyrGlySer 106  
 Db 606 ATAGAAAGGAAGCAAAATGGTGGAGGTCGGATATATGTTGGAATGAATATGCGAGC 665  
 Qy 107 TyrMetGluGluLysHisIleProProProAsnMetThrThrAsnGluArgValIle 126  
 Db 666 TACATAGAAGAAACACATTTCTCCCAACATGACCAACATGAGAGAGAGTGATT 725  
 Qy 127 ValProAlaAspProThrLeuTrpSerThrAspHisValArgGlnTrpLeuGluTrpAla 146

Db 726 GTACCTGCTGACCCCTACTTTTGTGAGCACAGATCATGTCAACAGATGGCTAGATGGCA 785  
 Qy 147 VallysGluTyrGlyLeuProAspValAspIleLeuLeuPheGlnAsnIleAspGlyLys 166  
 Db 786 ATAAAGNATACGGCTTCCAGATGTGATGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 845  
 Qy 167 GluLeuCysLysMetThrLysAspAspPheGlnArgLeuThrProSerTyrAsnAlaAsp 186  
 Db 846 GAATTTATGCAAGATGACCAAGAAGATTTCATGCTGACTCAGCGCAAGCTACAATGCTGAT 905  
 Qy 187 IleLeuLeuSerHisLeuHisTyrLeuArgGluArg 198  
 Db 906 ATCTTTCTCTCGCATCTTCACTTACTCTAGAGAAACTCCTCTTCCACATTTGACTTCAGAT 965  
 Qy 199 -----Gly 199  
 Db 966 GATGTTGATAAGGCTTACAAAACCTCCACGGTTAATGTCATGTAGAAATACAGGAGGT 1025  
 Qy 200 AlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnAlaThrThr 219  
 Db 1026 GCCTCTTTTATTTTCCCAACTCATCCGTTTATCAGAGGCAACCAAGANTCCCCAGC 1085  
 Qy 220 ArgProAspLeuProTyrGluGlnAlaArgSerAlaTyrThrSerHisSerHisPro 239  
 Db 1086 AGGCAAGATTTATCATATGAGCCCTCAAGAGATCTGCTTGGACAAATCATCTCGACCA 1145  
 Qy 240 ThrGlnSerLysAlaThrGlnProSerSerSerThrValProLysThrGluAspGlnArg 259  
 Db 1146 ---CCATCAAAAGCATCTCAACCA---TCTACCAAGTTCCTCAAAACAGAGAGCCCCAGG 1199  
 Qy 260 ProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGly 279  
 Db 1200 CCACAGCTAGATCTTATCAGATCTTGGGCAACCAAGCAGCGCATTTGCAATTCAGCA 1259  
 Qy 280 SerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeuSerSerSerSerSerSer 299  
 Db 1260 AGTGGGAGATTCAGCTCTGCAATCTTATGGAACCTGCTTTCGATAGTTCCTCACTCC 1319  
 Qy 300 AsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspProAspGluVal 319  
 Db 1320 AACTGCATTTCTTGGGAAGAACCAATGGAGAAATTAAGATGACCGATCCCGATGAAGTG 1379  
 Qy 320 AlaArgArgTrpGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArg 339  
 Db 1380 GCAAGACGTTGGGGGGAGAGAAAGCAACCCCAACATGAATGATGCAAGCTACCGCT 1439  
 Qy 340 AlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAla 359  
 Db 1440 GCACCTTCGTTACTACTACGATAAAATATATGACTAAAGTCCATGGCAAGCGCTATGCT 1499  
 Qy 360 TyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProProGluSerSer 379  
 Db 1500 TACAAATTTGATTTTTCATGGGATTCCTCAGGCTCTTCAACCTCATCTCCAGATCAACC 1559  
 Qy 380 MetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisProGlnLys 399  
 Db 1560 ATGTACAAGTATCCCTCAGAACTCCATATATGACTCATACCATGCAATCCACAGAG 1619  
 Qy 400 MetAsnPheValAlaProHisProProAlaLeuProValThrSerSerSerPhePheAla 419  
 Db 1620 ATGAATTTGTTGCCCCCATCCCGCAGCTTTTACCTGTGACATCGTCAAGTTCTTTTGA 1679  
 Qy 420 AlaProAsnProTyrTyrAsnSerProThrGlyIleTyrProAsnThrArgLeuPro 439  
 Db 1680 GCCCTTAATGCAATCTGGAATTCACCAACTGGAATTTATTCCAATACTCCGCTGCCA 1739  
 Qy 440 AlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451  
 Db 1740 GCTAGCCATATGCTCTCTCATCTTGGAACTTACTAT 1775

RESULT 13  
 AB073078

LOCUS AB073078 2131 bp mRNA linear ROD 17-OCT-2001  
 DEFINITION Mus musculus Erg mRNA, mouse homolog of Human ets-related gene ERG,  
 transcript variant 1, complete cds.  
 AB073078  
 ACCSSION AB073078.1 GI:16191716  
 VERSION 1  
 KEYWORDS Mus musculus (house mouse)  
 SOURCE Mus musculus  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1  
 REFERENCE Ozawa,R., Noguchi,H., Taylor,T.D., Takeda,T., Hattori,M. and  
 AUTHORS Sakaki,Y.  
 TITLE Mus musculus Erg mRNA  
 JOURNAL Published Only in Database (2001)  
 REFERENCE 2 (bases 1 to 2131)  
 AUTHORS Ozawa,R., Noguchi,H., Taylor,T.D., Takeda,T., Hattori,M. and  
 Sakaki,Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-OCT-2001) Masahira Hattori, The Institute of Physical  
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
 1-7-22 Suehiro-chou, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
 (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/  
 Tel:81-45-503-9111, Fax:81-45-503-9170)  
 FEATURES  
 source  
 1. 2131  
 /location=Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /chromosome="16"  
 /clone="Erg-1"  
 1. 2131  
 /gene="Erg"  
 148..1536  
 /gene="Erg"  
 /note="homolog of human ets-related gene (ERG), transcript  
 variant 1"  
 /codon\_start=1  
 /protein\_id="BAB69948.1"  
 /db\_xref="GI:16191717"  
 /translation="MIGTVDPDPAARHKEALSVVSEDSQSLFECAVGTGPHLAKETMTASS  
 SSDYQTSKMSRPVQDDWISQAPARVIRKCNPNQVNGSRNPPDCSVNKGKMGV  
 SPDTVNGSYGYSVEKEKVPFPNMTNERRVIVPADPLWSTDRVOWLEWAVKEVGLL  
 DVDFLLQNDIGELCKMTKDDFORLTPTS VNADILLSHLYRETPLPLTSDVDVKA  
 LQNSPLRHARNLDLPEPRRRAWGHSHLTPQSKAQPSPSAVPEKTDQRPQLDPY  
 QILGPTSRRLANPGSGIQLWQFLELLSDSSNCLTWEGTNGERFAMDPDPEVARRW  
 GERKSPNMYDKLRLRYDYDKNIMTKVHGKRYAKFDHGIAGIQLQPHPEPSLY  
 KPSDLFPMGYSYAHAPQKXNFVSPHPALPVTSSFFASNPYNSPTGGIVENTRLP  
 ASHMPSHLGTYY"

ORIGIN

Alignment Scores:  
 Pred. No.: 1.17e-141 Length: 2131  
 Score: 2151.00 Matches: 402  
 Percent Similarity: 91.07% Conservative: 16  
 Best Local Similarity: 87.58% Mismatches: 27  
 Query Match: 88.01% Indels: 14  
 DB: 10 Gaps: 4  
 US-09-902-772-2 (1-451) x AB073078 (1-2131)  
 QY 2 AlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGluCys 21  
 DB 172 GCAGCTCATATTAGGAGCCCTTGTAGTTGTGAGCGAGGACCACTACTATTAGTGT 231  
 QY 22 AlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGluTyr 41  
 DB 232 GCCTACGGAACGCCACACCTCGCTAAGACAGAGATGACCGCATCTCTTCCAGTACTAT 291  
 QY 42 GlyGlnThrSerLysMetSerProArgValProGlnGlnAspTTPLeuSerGlnProPro 61  
 DB 292 GCCCAGACATCCAGATGAGTCCAGAGTCCCTCAGCAGGACTGGCTGTCTCAGGCCCA 351

62 AlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSer 81  
 DB 352 GCCAGGGTCAACATCAAGATGAGTGCACCTTAGTCAAGTGAATGGTTCAGGAACTCA 411  
 QY 82 ProAspAspCysSerValAlaLysGlyLysMetValSerSerSerSerAspAsnValGly 101  
 DB 412 CCTGATGAGTGCAGTGTGAACAAAGTGGGAGATGGTGGCAGCCCGACTACTGTGGG 471  
 QY 102 MetAsnTyrGlySerTyrMetGluGluLysHisIleProProAsnMetThrThrAsn 121  
 DB 472 ATGAGCTACGGCAGCTACATGAGGAGAGAGATGGCCCTCCCAATATGACCAAAAT 531  
 QY 122 GluArgValIleValProAlaAspProThrLeuTyrSerThrAspHisValArgGln 141  
 DB 532 GAGCCGACAGATGATCGTCCCTGCAGATCTACTCTGTGGAGCAGACCATGTCTCGACAG 591  
 QY 142 TrpLeuGluTyrAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPheGln 161  
 DB 592 TGGCTGGAGTGGCGCGTGAAGAATATGCCCTCTCTCGATGGAGCTTACTATTATTCAG 651  
 QY 162 AsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThrPro 181  
 DB 652 AATATCGATGGAGAGAGCTGTGCAAGATGACAAAGGATGACTTCCAGCGGCTCAGCGG 711  
 QY 182 SerTyrAsnAlaAspIleLeuLeuSerHisIleHisTyrLeuArgGluArgGlyAlaThr 201  
 DB 712 AGCTACAATGCGCAGATCTTCTCTCATCTCCACTACCTCAGAGAGACTCCCTTCCA 771  
 QY 202 PheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrThrArgPro 221  
 DB 772 CATCTG-----ACTTCC-----GATGACGTTGATAAGGCTTTACAAAACCTCCA 816  
 QY 222 -----AspLeuProTyrGluGlnAlaArgArgSerAlaTyr 233  
 DB 817 CGGTAAATGTCATGCCAGAAACACAGATTTACCTATGAGCCTCCAGAGATCACCGTGG 876  
 QY 234 ThrSerHisSerHisProThr---GlnSerLysAlaThrGlnProSerSerSerThrVal 252  
 DB 877 ACCGCCACAGCCACTCACCCCTCAGTCCAAAGTGTCTCAGCCATCTCCCTCTCAGTGT 936  
 QY 253 ProLysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSer 272  
 DB 937 CCAAAAACCTGAAGACAGCGTCTCAGTTAGTATCTTACCATAGTCTGGACCCAGT 996  
 QY 273 SerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTyrGlnPheLeuLeuGluLeu 292  
 DB 997 AGCCGCTTGTCTAATCCAGTAGTGGCAGATCCAGCTGTGGCAGTCTCTGTCTGACTC 1056  
 QY 293 LeuSerAspSerSerAsnSerAsnCysIleThrTyrGluGlyThrAsnGlyGluPheLys 312  
 DB 1057 CTGTTCAGACAGCTCCAACTCCAACTCCATCACCTGGGAAGGCACCAACGGGGAGTTCAAG 1116  
 QY 313 MetThrAspProAspGluValAlaAlaArgArgTyrGlyGluArgLysSerLysProAsnMet 332  
 DB 1117 ATGACAGACCCGAGCGAGTGGCTCGCGCTGGGGGAGAGAGAGAGAGAGAGAGAGAG 1176  
 QY 333 AsnTyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLys 352  
 DB 1177 AACTATGACAAGCTCAGCGCGCCCTCCGCTACTACTACGACAAACATCATGACCAAG 1236  
 QY 353 ValHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGln 372  
 DB 1237 GTGCACGGGAAGCGGTACCGCTACAGTTGTACTTCCAGGGATTTGCCAGGCCCTTGCAG 1296  
 QY 373 ProHisProProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSer 392  
 DB 1297 CCCCACCTCTCTGAGTCTCGTCCCTGTACAAGTACCCCTCCGACCTGCCATACATGGCTCC 1356  
 QY 393 TyrHisAlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProVal 412  
 DB 1357 TATCAGCCCAACCCAGAGATGAACCTTGTGTCTCCCAACCTCCCGCTCTCCAGTCTC 1416  
 QY 413 ThrSerSerSerPhePheAlaProAsnProTyrTyrAsnSerProThrGlyGlyIle 432

Db	1417	ACATCTTCAGTTCTTTGGTCTCCCGAACCCATACCTGGAAATTCACCGACTGGGGGCATC	1476
Qy	433	TyrProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr	451
Db	1477	TACCCGAACACTAGGTCTCCCGACGACCATATGCCCTCTCACCTGGSCACTACTAC	1533
RESULT 14			
AX657230			
LOCUS	AX657230	1291 bp	DNA
DEFINITION	Sequence 3 from Patent WO03000724.		linear
ACCESSION	AX657230		
VERSION	AX657230.1	GI:29160008	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
Karyotype	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1		
AUTHORS	Luyten, F., de Bari, C. and Dell'Accio, F.		
TITLE	Polynucleotide sequences and vectors useful for the prevention or		
	treatment of bone or cartilage-related disorders		
JOURNAL	Patent: WO 0300724-A 3 03-JAN-2003;		
	Tigenix N.V. (BE)		
FEATURES	Location/Qualifiers		
source	1..1291		
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
CDS	146..1228		
	/note="unnamed protein product"		
	/codon_start=1		
	/protein_id="CAD79712.1"		
	/db_xref="GI:29160009"		
	/db_xref="RETR:EMBL.CAD79712"		
	/translation="MVGSPDTVMGVGSMEEKMPDPMTNTRRVIVPADTFLNST		
	DHVRQMLEAVKVGJFDPVNNILFQNIQDKELCKMTKDDQRLTPPSNADILSHLHY		
	LRERGAATFPNTSVYPEATQRTITPDLPEYEPERSAMTGHPHTQSKAAQSPST		
	EFVTEQRPQLDPYQLGTSRLANPGSQIQLWQFLLELLSDSNSSCITWEGTNG		
	YKPTWDPEVARWGWRKSPNNYDKLSRALRYVDKNTMTKVHGKRYAYKPDFHGI		
	AQALPHPPSSILVKYPSDLPLVYMGSVYHAHPQKNFVAPHPALPVTSSSFFAAAPNPV		
	NSTGGIYPNTRLPSTSEMPSHLGTYY"		
ORIGIN			
Alignment Scores:			
Pred. No.:	3.35e-128	Length:	1291
Score:	1955.50	Matches:	356
Percent Similarity:	96.53%	Conservative:	6
Best Local Similarity:	94.93%	Mismatches:	12
Query Match:	80.01%	Indels:	1
DB:	6	Gaps:	1
US-09-902-772-2 (1-451) x AX657230 (1-1291)			
Qy	78	SerArgAsnSerProAspAspCysSerValAlaLysGlyGlyLysMetValSerSer	97
Db	101	TCAAGGAACCTCTCTCTGAATGATGAGTGTGGCCNAGCGGGAAGATGTGGCGACGCCA	160
Qy	98	AspAsnValGlyMetAsnTyrGlySerTyrMetGluLysHisLileProProProAsn	117
Db	161	GACACCGTTGGGATGAACCTACGGCAGCTACATGAGGAGAGACATGCCACCCCAAC	220
Qy	118	MetThrThrAsnGluArgArgValilleValProAlaAspProThrLeuTyrSerThrAsp	137
Db	221	ATGACCACGACACGACGACGAGTATTGTCGACGACGATCCTACGCTATGGAGTACAGAC	280
Qy	138	HisValArgGlnTyrLeuGluTyrAlaValLysGluTyrGlyLeuProAspValAspile	157
Db	281	CAITGTCGGCAGTGGCTGGAGTGGCGGTGAAGAATATGGCCCTCCACAGCTCAACATC	340
Qy	158	LeuLeuPheGlnAenilleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGln	177
Db	341	TTGTATTTCAGAACATCGATGGGAAGAACTGTGCAAGATGACCAAGAGACGATTCACG	400

Tigenix N.V. (BE)		Location/Qualifiers	
FEATURES	source	1. .1372	
		/organism="Homo sapiens"	
		/mol_type="unassigned DNA"	
		/db_xref="taxon:9606"	
CDS		146. .1309	
		/note="unnamed protein product"	
		/codon_start=1	
		/protein_id="CAD79714.1"	
		/db_xref="GI:29160013"	
		/db_xref="EMBL:CAD79714"	
		/translation="MVGSPDTGVGNYSYMEKHPNNTNERRVIVPADPLAKST DHVQLEWAWKETGLDPAVILLFQNDIGKELCKMTKDDFRLTPSTNADILLSHLHY LREPLHLSDDDVDKALQNSPRLMHARTNGGAFFPNTSVYPEATQRIITTPDLPY EPPRSWHTGHPTPOSKAQSPFSPKTEPQDPQLDPYQILGPTSSRLANPGSGQ IQLWQFLLELLSDNSSTWECTWEGTFKMTDPDEVAREWGERKSKPNMYDKLARA LRVYDKNINTKVGKAYKVFDPHQAQALQHPPESSLYKYPSDLPYMGSVHAHPQ KMFVAPHPALPYTSSFFPAAPNPYNSFTGGIYPNTRLPSTHMPSHLGTYY"	
ORIGIN			
Alignment Scores:		Length: 1372	
Pred. No.:	4.23e-126	Matches: 355	
Score:	1926.00	Conservative: 6	
Percent Similarity:	89.80%	Mismatches: 13	
Best Local Similarity:	88.31%	Indels: 28	
Query Match:	78.81%	Gaps: 2	
DB:	6		
US-09-902-772-2 (1-451) x AX657234 (1-1372)			
QY	78	SerArgAsnSerProAspAspCysSerValAlaLysGlyGlyLysMetValSerSerSer	97
DB	101	TCAAGGAATCTCTCTGATGAATGAGTGTGCCAAGCGGGAAGATGGTGGCGAGCCCA	160
QY	98	AspAsnValGlyMetAsnTyrGlySerTyrMetGluGluLysHisIleProProAsn	117
DB	161	GACACCGTTGGATGAACACTACGGCAGCTACATGGAGGAGAGACATGCCACCCCAAC	220
QY	118	MetThrThrAsnGluArgValIleValProAlaAspProThrLeuTrpSerThrAsp	137
DB	221	ATGACCAAGCGGCGAGAGTATATGTCGACAGATCTTACGCTATGGAGTACAGAC	280
QY	138	HisValArgGlnTrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIle	157
DB	281	CATGTGGCGCAGTGCCTGAGTGGCGGTGAAGATATGCGCTTCCAGACGTCACATC	340
QY	158	LeuLeuPheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGln	177
DB	341	TTGTTATTCCAGAACATCGATGGGAAGGAATGTGCAAGATGACCAAGGACGACTTCCAG	400
QY	178	ArgLeuThrProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGlu	197
DB	401	AGGCTCACCCCCAGCTACACGCGGACATCTTCTCTCACATCTCCACTACCTCAGAGAG	460
QY	198	Arg-----	198
DB	461	ACTCCTCTCCACATTTGACTTCAGATGATGTTGATAAGCCTTACAAAACTCTCCACGG	520
QY	199	-----GlyAlaThrPheIlePheProAsnThrSerValTyr	210
DB	521	TTAATGCATGCTAGNAACACAGGGGGTGCAGCTTTATTTTCCCAATACTTTCAGTATAT	580
QY	211	ProGluAlaThrGlnArgIleThrThrArgProAspLeuProTyrGluGlnAlaArgArg	230
DB	581	CCTGAAGCTACGCAAGAAATTAACACTAGGCGAGATTACCATATGAGCCCCCAGGAGA	640
QY	231	SerAlaTrpThrSerHisSerHisProThr---GlnSerLysAlaThrGlnProSerSer	249
DB	641	TCAGCCTGGACCGGTACGCGGCCACCCCGCCAGTCCGAAGCTGTCAACCATCTCCT	700
QY	250	SerThrValProLysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGly	269

Search completed: May 27, 2004, 22:55:53  
Job time : 5643 secs

Db	701	TCCACAGTCCCAAACTGAACACCAGCGTCTCTAGTTAGATCCTTATCAGATTCTTGA	760
QY	270	ProThrSerSerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeu	289
DB	761	CCAAACAAGTAGCGCTTGCAATCCAGGCACGTGGCCAGATCCAGCTTGGCAGTTCCTC	820
QY	290	LeuGluLeuLeuSerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGly	309
DB	821	CTGGAGCTCCTCTCGACAGCTCCAACTCCAGCTGCATCACCCTGGGAAGGCCAACGGG	880
QY	310	GluPheLysMetThrAspProAspGluValAlaIaAaGtGtGtGtGtGtGtGtGtGtGtGt	329
DB	881	GAGTTCAGATCAGGATCCCGACGAGGTGGCCGCGCTGGGGAGAGCGGAAGACAAA	940
QY	330	ProAsnMetAsnTyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIle	349
DB	941	CCCAACATGAATACGATAAGCTCAGCGCGCCCTCCGTTACTACTATGACAAGAACATC	1000
QY	350	MetThrLysValHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGln	369
DB	1001	ATGACCAAGGTCCATGGGAGGCGCTACGCTACAAGTTCGACTTCCACGGGATCGCCAG	1060
QY	370	AlaLeuGlnProHisProGluSerSerMetTyrLysTyrProSerAspLeuProTyr	389
DB	1061	GCCTTCAGCCCGACCCCGCGAGTCACTCTGTACAAGTACCCCTCAGACCTCCCGTAC	1120
QY	390	MetSerSerTyrHisAlaHisProGlnLysMetAsnPheValAlaProHisProProAla	409
DB	1121	ATGGGCTCTATCACGCCCCACAGAGATGAATTTTGTGGCGCCGCCCTCCAGGC	1180
QY	410	LeuProValThrSerSerSerPheAlaAlaProAsnProTyrTrpAsnSerProThr	429
DB	1181	CTCCCGTGCATCTTCCAGTTTTTTTGTGCCCCAACCCATACCTGGAATTCACCACT	1240
QY	430	GlyGlyIleTyrProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThr	449
DB	1241	GGGGGTATATACCCCAACACTAGGCTCCCAACAGCCATATGCTTCTCATCTGGGCACT	1300
QY	450	TyrTyr 451	
DB	1301	TACTAC 1306	

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 27, 2004, 19:25:13 ; Search time 610 Seconds  
(without alignments)  
3140.881 Million cell updates/sec

Title: US-09-902-772-2

Perfect score: 2444

Sequence: 1 MASTIKALSVSDQSLFE.....IYENTRLPAHMPHSLGTYV 451

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Dlopp 6.0 , Delext 7.0

Searched: 337363 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+\_B2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_EPOOL/US09902772/runat\_27052004\_162840\_2136/app\_query.fasta\_1.647  
-DB=N Geneseq\_29Jan04 -QMT=fastcap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0  
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFWT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USERS=US09902772 -CGEN 1 1 470 @runat\_27052004\_162840\_2136 -NCPU=6 -ICPU=3  
-NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREAS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_29Jan04.\*

1: geneseq1980s.\*  
2: geneseq1990s.\*  
3: geneseq2000s.\*  
4: geneseq2001as.\*  
5: geneseq2001bs.\*  
6: geneseq2002as.\*  
7: geneseq2003as.\*  
8: geneseq2003bs.\*  
9: geneseq2003cs.\*  
10: geneseq2004as.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2444	100.0	1447	2 AAX26551	Aax26551 DNA encod
2	2414.5	98.8	1528	2 AAX26552	Aax26552 DNA encod
3	2327.5	95.2	1509	8 ABZ68766	Abz68766 Nucleotid
4	2182	89.3	1389	7 ABX08759	Abx08759 Angiogene
5	2182	89.3	3166	7 ABX76295	Abx76295 Lung can
6	2182	89.3	3166	9 ADB75274	Adb75274 Prostata
7	1955.5	80.0	1291	8 ABZ68767	Abz68767 Nucleotid
8	1926	78.8	1372	8 ABZ68769	Abz68769 Nucleotid

9	1807.5	74.0	1219	8	ABZ68768	Abz68768 Nucleotid
10	1624.5	66.5	1359	8	ADA02692	Ada02692 Human Fli
11	1624.5	66.5	1359	9	ADB72430	Adb72430 Human Fli
12	1624.5	66.5	2938	2	AAQ50644	Aaq50644 Human Hum
13	1624.5	66.5	2957	6	ABV94304	Abv94304 Breast ca
14	1624.5	66.5	2957	8	ABK84139	Abk84139 Human CON
15	1624.5	66.5	2957	8	ADA02691	Ada02691 Human Fli
16	1624.5	66.5	2957	9	ADB72429	Adb72429 Human Fli
17	1613.5	66.0	1359	8	ADA02689	Ada02689 Mouse Fli
18	1613.5	66.0	1359	9	ADB72427	Adb72427 Mouse Fli
19	1613.5	66.0	1729	8	ADA02688	Ada02688 Mouse Fli
20	1613.5	66.0	1729	9	ADB72426	Adb72426 Mouse Fli
21	1599	65.4	2954	4	AAH02915	Aah02915 Human she
22	1506.5	61.6	1890	2	AAQ50662	Aaq50662 Human Hum
23	965	39.5	1431	6	ABZ73253	Abz73253 DNA encod
24	948	38.8	567	4	AAI20932	Aai20932 Probe #10
25	948	38.8	567	4	ABA66003	Abas66003 Human foe
26	948	38.8	567	4	AAI46176	Aai46176 Probe #14
27	948	38.8	567	4	ABA48124	Abas48124 Human bre
28	948	38.8	567	4	ABA33090	Abas33090 Probe #11
29	948	38.8	567	4	AAK40160	Aak40160 Human bon
30	948	38.8	567	4	AAK14426	Aak14426 Human bra
31	948	38.8	567	4	ABS39742	Abas39742 Human liv
32	948	38.8	567	5	AAI06643	Aai06643 Probe #6
33	948	38.8	567	6	ABS14215	Abas14215 Human gen
34	863	35.3	473	4	ABA68128	Abas68128 Human foe
35	863	35.3	473	4	AAI48339	Aai48339 Probe #17
36	863	35.3	473	4	ABA50193	Abas50193 Human bre
37	863	35.3	473	4	ABA35152	Abas35152 Probe #13
38	863	35.3	473	4	AAK42265	Aak42265 Human bon
39	863	35.3	473	4	AAK16512	Aak16512 Human bra
40	863	35.3	473	5	ABS41873	Abas41873 Human liv
41	863	35.3	473	5	AAI08696	Aai08696 Probe #86
42	833.5	34.1	73725	8	ADA02690	Ada02690 Human Fli
43	833.5	34.1	73725	9	ADB72428	Adb72428 Human Fli
44	813.5	33.3	72732	8	ADA02687	Ada02687 Mouse Fli
45	813.5	33.3	72732	9	ADB72425	Adb72425 Mouse Fli

#### ALIGNMENTS

RESULT 1

AXX26551

ID AAX26551 standard; DNA; 1447 BP.

XX AAX26551;

XX 14-JUN-1999 (first entry)

XX DNA encoding chicken C-11 protein.

XX Chicken; C-11 protein; cell calcification inhibiting activity;

XX cell calcification inhibiting agent; c-erg protein; arthritis deformans;

XX ossification; spinal column ligament; ss.

XX Gallus sp.

XX Key Location/Qualifiers

FT CDS 63..1418

FT /\*tag= a

XX JP11075871-A.

XX 23-MAR-1999.

XX 29-MAY-1998; 98JP-00166076.

XX 18-JUN-1997; 97US-00878177.

XX 20-JUN-1997; 97US-0050297P.

XX (CHUS ) CHUGAI PHARM CO LTD.

XX (UTYPE-) UNIV PENNSYLVANIA.



DR WFI; 1998-257708/22.  
 DR P-PSDB; AAY01520.  
 XX An active protein for inhibiting cell calcification - useful for  
 PT measuring the calcification of a cell, for diagnosing arthritis deformans  
 PT or ossification of spinal column ligament.  
 XX Disclosure; Page 7-8; 15pp; Japanese.  
 XX The present sequence encodes a chicken C-11 protein which has cell  
 CC calcification inhibiting activity. The specification also describes a  
 CC cell calcification inhibiting agent containing c-erg protein (AAY01521).  
 CC The proteins are used for measuring the calcification of a cell, for  
 CC diagnosing arthritis deformans or ossification of spinal column ligament  
 XX Sequence 1447 BP; 440 A; 374 C; 317 G; 316 T; 0 U; 0 Other;  
 SQ

Alignment Scores: 2.01e-199 Length: 1447  
 Pred. No.: 2444.00 Matches: 451  
 Score: 2444.00 Conservative: 0  
 Percent Similarity: 100.00% Mismatches: 0  
 Best Local Similarity: 100.00% Indels: 0  
 Query Match: 100.00% Gaps: 0  
 DB: 2

US-09-902-772-2 (1-451) x AAX26551 (1-1447)

Qy 1 MetLaserThrIleYsGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20  
 Db 63 ATGCAGAGCACTATTAGGAAGCAATTATCAGTGTGTAGTGAACCGACGCTTTGTTGAG 122  
 Qy 21 CysAlaTyrGlySerProHisLeuAlaIleThrGluMetThrAlaSerSerSerGlu 40  
 Db 123 TGTGCTACGGATCGCCCACTTGTCAAGACAGCAAGAAATGACAGCTCTCTCCAGTGAA 182  
 Qy 41 TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnPro 60  
 Db 183 TATGGCAAAATCAATCAAGATGAGCCCGCGCTTCCCGAGGAGCTGTTATCACAGCCC 242  
 Qy 61 ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80  
 Db 243 CGGCCAGAGTTACATTAAGATGGAGTGTAACCCAAACAGGTTAATGGTCAAGGAAT 302  
 Qy 81 SerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsnVal 100  
 Db 303 TCACCTGATGACTCGCGTGGCAAAAGAGGGAATATGGTTAGCAGTTGACAGCAATGTT 362  
 Qy 101 GlyMetAsnTyrGlySerTyrMetGluGlyLysHisIleProProAsnMetThrThr 120  
 Db 363 GGGATGAACATATGAGAGCTACATGGAGAGAGAGCATATTCGCTCCCAATATGACAAAC 422  
 Qy 121 AsnGluArgArgValIleValProAlaAspProThrLeuTyrSerThrAspHisValArg 140  
 Db 423 AATGAACAGAGATTATTGTGCCAGCAGATCCTACGTTATGGAGCACAGACCATGTACGG 482  
 Qy 141 GlnTyrLeuGluTyrAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPhe 160  
 Db 483 CAGTGGCTGGAGTGGGAGTGAAGAGATGCTTTCAGACGCTGGACATCTTGTCTTC 542  
 Qy 161 GlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThr 180  
 Db 543 CAGAACATTGATGGGAAGAGTTGTGTAAATGACCAAAAGATGACTTCCAGAGACTCACG 602  
 Qy 181 ProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArgGlyVala 200  
 Db 603 CCGAGCTATACGCAGATATCTCTGTCAACCTACCTACCTACCTACCTACCTACCTACCTAC 662  
 Qy 201 ThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrArg 220  
 Db 663 ACTTTTATTTTCCAAATACATCACTTTACCCAGAGCAACGCAAGGAATTAACAACAGG 722  
 Qy 221 ProAspLeuProTyrGluGlnAlaArgSerAlaThrThrSerHisSerHisProThr 240

Db 723 CCAGATTTACCTTATGAGCAAGCGAGGAGATCAGCGTGGAGCTCAGCCATCCCACT 782  
 Qy 241 GlnSerLysAlaThrGlnProSerSerSerThrValProLysThrGluAspGlnArgPro 260  
 Db 783 CAGTCAAAAGCTACCAACCATCATCTTCAACAGTGGCCAAACAGAAAGACACCGCTCT 842  
 Qy 261 GlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGlySer 280  
 Db 843 CAGTTAGATCTTATCAGATTCTTTGGACCGACAGCAGCGCTCTTGAATTCAGAGGAGT 902  
 Qy 281 GlyGlnIleGlnLeuTyrGlnPheLeuLeuLeuLeuSerSerSerSerSerSerAsn 300  
 Db 903 GGGCAGATACAGCTATGCGAGTTCCTACTGAGCTTCTTGGACAGCTTCAACTCCCAAC 962  
 Qy 301 CysIleThrTyrGluGlyThrAsnGlyGluPheLysMetThrAspProAspGluValAla 320  
 Db 963 TGCATCACCTGGAGGGGACAAATGGGGAGTTCAAGATGACAGACCCCTGATGAAGTGGCT 1022  
 Qy 321 ArgArgTyrGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArgAla 340  
 Db 1023 CGGCGTTGGGAGAGAGAGAAACCTTAACATGAACATATGACAACTCAGCCGTGCA 1082  
 Qy 341 LeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaTyr 360  
 Db 1083 CTTCGCTACTACTATGACAAATAATATATGACTTAAGTTCATGTTAAACGCTATGCTAC 1142  
 Qy 361 LysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGluSerSerMet 380  
 Db 1143 AAATTTGATTTCCACGGAATCGCTCAGGCCCTCCAGCCTCACCCTCCAGAAATCATCCATG 1202  
 Qy 381 TyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisProGlnLysMet 400  
 Db 1203 TACAAATACCATCCAGCTCCCTCATGATGATGCTTCCATACCATGCACACCCCGAGAGATG 1262  
 Qy 401 AsnPheValAlaProHisProAlaLeuProValThrSerSerSerPhePheAlaAla 420  
 Db 1263 AACTTTGTAGTCCCATCCCTCTGCTTTGCCCGTAACCTCATCCAGCTTTTGTCTGCC 1322  
 Qy 421 ProAsnProTyrTyrAsnSerProThrGlyGlyIleTyrProAsnThrArgLeuProAla 440  
 Db 1323 CCTAATCCATACCTGGAATTCACCAACTGGAGCATCTACCCCAATACCGAGCTGCCAGCT 1382  
 Qy 441 AlaHisMetProSerHisLeuGlyThrTyrTyr 451  
 Db 1383 GCTCATATGCTTCCCATCTTGGCACCTACTAC 1415

RESULT 2  
 AAX26552  
 ID AAX26552 standard; DNA; 1528 BP.  
 XX AAX26552;  
 AC AAX26552;  
 XX 14-JUN-1999 (first entry)  
 DT DNA encoding chicken c-erg protein.  
 DE  
 XX Chicken; C-11 protein; cell calcification inhibiting activity;  
 KW cell calcification inhibiting agent; c-erg protein; arthritis deformans;  
 KW ossification; spinal column ligament; ss.  
 OS Gallus sp.  
 XX Key Location/Qualifiers  
 FH 63..1499  
 FT /\*tag= a  
 PN JP11075871-A.  
 XX 23-MAR-1999.  
 XX 29-MAY-1998; 98JP-00166076.  
 XX 18-JUN-1997; 97US-00878177.

PR 20-JUN-1997; 97US-0050297P.  
 XX (CHUS ) CHUGAI PHARM CO LTD.  
 PA (UYPE-) UNIV PENNSYLVANIA.  
 XX WPI; 1999-257708/22.  
 DR P-PSDB; AAY01521.  
 XX  
 XX An active protein for inhibiting cell calcification - useful for  
 PT measuring the calcification of a cell, for diagnosing arthritis deformans  
 PT or ossification of spinal column ligament.  
 XX  
 XX Disclosure; Page 8-9; 15pp; Japanese.  
 XX  
 XX The present sequence encodes a chicken c-erg protein. The specification  
 CC also describes a chicken C-11 protein (AAY01520) which has cell  
 CC calcification inhibiting activity and a cell calcification inhibiting  
 CC agent containing c-erg protein. The proteins are used for measuring the  
 CC calcification of a cell, for diagnosing arthritis deformans or  
 CC ossification of spinal column ligament  
 XX  
 XX Sequence 1528 BP; 464 A; 395 C; 330 G; 339 T; 0 U; 0 Other;  
 SQ

Alignment Scores:  
 Pred. No.: 7, 26e-197 Length: 1528  
 Score: 2414.50 Matches: 450  
 Percent Similarity: 94.14% Conservatives: 0  
 Best Local Similarity: 94.14% Mismatches: 1  
 Query Match: 98.79% Indels: 27  
 DB: 2 Gaps: 1

US-09-902-772-2 (1-451) x AAX26552 (1-1528)

QY 1 MetAlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20  
 Db 63 ATGGCAAGCACTATTATAGGAAGCACTTATCAGTGGTGAAGTGAAGCAAGTCTCTTTTGAG 122  
 QY 21 CysAlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerSerGlu 40  
 Db 123 TGTGCTTCAGGATCCGCCACCTTCGCAAGACAGAAATGACAGCTCTCTCTCCAGTAA 182  
 QY 41 TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnPro 60  
 Db 183 TATGGCAAAACATCAAGATGAGCCGCGGTCCCAAGCAGGACTGGTTATCAAGCC 242  
 QY 61 ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80  
 Db 243 CCGGCCAGAGTTACCATTAAGATGAGGTGAACCCAAACCCAGTTATGGTCAAGGAT 302  
 QY 81 SerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerSerAspAsnVal 100  
 Db 303 TCACCTGATGACTGCAGCGTGGCAAAAGAGGGAATAATGGTTAGCAGTTCAGACAATGTT 362  
 QY 101 GlyMetAsnTyrGlySerTyrMetGluGlnLysHisIleProProProAsnMetThrThr 120  
 Db 363 GGGATGAACATATGGAAGCTACATGGAAGAGAGCATATTCGCGCTCCAAATATGACAACC 422  
 QY 121 AsnGluArgValIleValProAlaAspProThrLeuTyrSerThrAspHisValArg 140  
 Db 423 AATGAACGAGAGTATTGTGTCAGCAGATCTTACGTTATGGAGCAGACCATGATCAGG 482  
 QY 141 GlnTyrLeuGluTyrAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPhe 160  
 Db 483 CAGTGGCTGGAGTGGGACGTGAAGGAGTATGCTCTTCAGAGCTGGACATCTTGTGTTC 542  
 QY 161 GlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThr 180  
 Db 543 CAGAACATATGATGGGAAGAGTGTGTAAATATGACCAAGATGATTCAGAGATCTCAG 602  
 QY 181 ProSerTyrAsnAlaAspIleLeuSerHisLeuHisTyrLeuArgGluArg----- 198  
 Db 603 CCGAGCTATAACGACATATCTCTCTGTACACCTACACTACCTCAGAGAGACTCTCTT 662

QY 198 ----- 198  
 Db 663 CCACATTTGACTTCCAGATGATGTTGATAAGGCTTACAAAACCTCTCCAGGTTAATGAT 722  
 QY 199 -----GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAla 213  
 Db 723 GCTAGAAACACAGGAGGAGCCACATTTATTTTCCAAATACATCAGTTTACCCAGAGCA 782  
 QY 214 ThrGlnArgIleThrThrArgProAspLeuProTyrGluGlnAlaArgSerSerAlaTyr 233  
 Db 783 ACGCAAGAATAAACCAAGCCAGATTACCTTATGAGCAAGCAGGAGATCAGCGTGG 842  
 QY 234 ThrSerHisSerHisProThrGlnSerLysAlaThrGlnProSerSerSerThrValPro 253  
 Db 843 ACGAGTCACAGGCATCCCACTCAGTCAAAAGTACCCCAACCATCATCTTCAACAGTGCC 902  
 QY 254 LysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSer 273  
 Db 903 AAAACAGAGACAGCGTCTCAGTTAGATCCTTATCAGATTCTTGGACCGCAGCAGCAG 962  
 QY 274 ArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTyrGlnPheLeuLeuLeuLeu 293  
 Db 963 CGTCTTTGCAATCCAGGGAGTGGGAGATACAGCTATGCGAGTTCTTACTGGAGCTTCTG 1022  
 QY 294 SerAspSerSerAsnSerAsnCysIleThrTyrGluGlyThrAsnGlyGluPheLysMet 313  
 Db 1023 TCGGACAGCTCCCACTCCCACTGCTCATCACCTGGGAGGCGACAAATGGGGAGTTCAGATG 1082  
 QY 314 ThrAspProAspGluValAlaArgTyrGlyGluArgLysSerLysProAsnMetAsn 333  
 Db 1083 ACAGACCTTGATGAAGTGGCTGGGAGTGGGAGAGGAAAGAAAGCAACTTACATGAAC 1142  
 QY 334 TyrAspLysLeuSerArgAlaLeuArgTyrTyrAspLysAsnIleMetThrLysVal 353  
 Db 1143 TATGACAAACTCAGCGTGCACCTCGCTACTACTATGACAAAATAATTATGACTAAAGTT 1202  
 QY 354 HisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnPro 373  
 Db 1203 CATGGTAAACGCTATGCTTACAAATTTGATTTCCAGGAATCGCTCAGCGCTCCAGCT 1262  
 QY 374 HisProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyr 393  
 Db 1263 CACCTCCAGATCATCCATGATGATCAATACCCATCAGACTCCCTTACATGAGTCTCTAC 1322  
 QY 394 HisAlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProValThr 413  
 Db 1323 CATGCAACACCCAGAGATGAACCTTGTAGCTCCCATCCCTCTCTTGGCCGTAACC 1382  
 QY 414 SerSerSerPhePheAlaAlaProAsnProTyrTyrAsnSerProThrGlyGlyIleTyr 433  
 Db 1383 TCATCCAGCTTTTGTGCTGCCCTTATCCATCTGGAATTCACCAACTGGAGCACTCTAC 1442  
 QY 434 ProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451  
 Db 1443 CCCAATACAGGCTGCCAGCTGCTCATATGCTTCCCTTCCCTTGGCACCTACTAC 1496

RESULT 3  
 ABZ68766 standard; DNA; 1509 BP.  
 ID ABZ68766  
 XX AC ABZ68766;  
 XX  
 XX 16-MAY-2003 (first entry)  
 DT  
 XX  
 DE Nucleotide sequence of human ERG splice variant C-1-1.  
 XX  
 KW Human; splice variant; ERG; Ets; transcription factor; C-1-1; C-1-2;  
 KW C-1-3; C-1-4; chondrocyte; articular chondrocyte; bone disease;  
 KW cartilage disease; tumour; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 FH



Db 1411 AGCCATATGCTTCTCATCTGGGCACTTACTAC 1443

RESULT 4  
 ABX08759  
 ID ABX08759 standard; cDNA; 1389 BP.  
 XX  
 AC ABX08759;  
 XX  
 DT 21-JAN-2003 (first entry)  
 XX  
 DE Angiogenesis-associated human polynucleotide sequence #21.  
 XX  
 DE Human; angiogenesis-associated transcript; angiogenesis;  
 KW angiogenesis-associated disease; cancer; cytostatic; gene therapy; gene;  
 KW ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200279492-A2.  
 XX  
 PD 10-OCT-2002.  
 XX  
 PF 14-FEB-2002; 2002WO-US004915.  
 XX  
 PR 14-FEB-2001; 2001US-00784356.  
 PR 22-FEB-2001; 2001US-00791390.  
 PR 19-APR-2001; 2001US-0285475P.  
 PR 03-AUG-2001; 2001US-0310025P.  
 PR 13-NOV-2001; 2001US-0350666P.  
 PR 29-NOV-2001; 2001US-0334244P.  
 XX  
 PA (EOSB-) EOS BIOTECHNOLOGY INC.  
 XX

Murray R, Glynn R, Watson SR, Aziz N;  
 WPI; 2003-040681/03.  
 DR P-PSDB; ABU03476.  
 XX  
 XX  
 PT Detecting angiogenesis-associated transcript in a cell for diagnosing and  
 PT treating cancer by contacting a sample with a polynucleotide that  
 PT exhibits changes in expression level as a function of time in tissue  
 PT undergoing angiogenesis.  
 XX  
 PS Example 2; Page 200; 291pp; English.  
 XX

CC The present invention relates to methods and compositions for detecting  
 CC an angiogenesis-associated transcript in a cell in a patient. The method  
 CC involves contacting a biological sample from the patient with a  
 CC polynucleotide that selectively hybridises to a sequence at least 80%  
 CC identical to any of the angiogenesis-associated human polynucleotide  
 CC sequences given in the specification. These angiogenesis-associated  
 CC polynucleotide sequences comprise genes that exhibit changes in  
 CC expression levels as a function of time in tissue undergoing  
 CC angiogenesis. The method and the polynucleotide sequences of the  
 CC invention are useful for diagnosing and treating angiogenesis and  
 CC angiogenesis-associated diseases e.g. cancer. The polynucleotide  
 CC sequences are also useful in the gene therapy of such disorders. The  
 CC angiogenesis-associated proteins encoded by the polynucleotide sequences  
 CC are useful as a vaccine for therapeutic and prophylactic immunisation.  
 CC ABX08735-ABX08853 represent angiogenesis-associated polynucleotide  
 CC sequences  
 XX

SQ Sequence 1389 BP; 359 A; 440 C; 323 G; 267 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 5,048-177 Length: 1389  
 Score: 2182.00 Matches: 407  
 Percent Similarity: 91.72% Conservative: 14  
 Best Local Similarity: 88.67% Mismatches: 24  
 Query Match: 89.28% Indels: 14  
 DB: 7 Gaps: 4

US-09-902-772-2 (1-451) x ABX08759 (1-1389)

QY	2	AlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGluCys	21
Db	25	GCAGCTCATATCAAGAAAGCCCTTATCAGTTGTAGTGAGGACCACTGCTGTTGTAGGT	84
QY	22	AlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerSerGluTyr	41
Db	85	GCCTACGGAAACGCCACACTGGCTAAGACAGAGATGACCGGTCTCTCTCCAGCGACTAT	144
QY	42	GlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnPro	61
Db	145	GGACAGACTTCCAGATGAGCCCGCTCCCTCAGCAGGATTGGCTGCTCAACCCCA	204
QY	62	AlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSer	81
Db	205	GCCAGGGTCACCATCAAAATGGAATGAACCTAGCCAGGTGAATGGCTCAAGCACTCT	264
QY	82	ProAspAspCysSerValAlaLysGlyLysMetValSerSerSerSerAspAsnValGly	101
Db	265	CCTGATGAATGCGAGTGTGCCAAGCGGGAAGATGGTGGCCAGCCAGACCGTTGGG	324
QY	102	MetAsnTyrGlySerTyrMetGluLysHisIleProProProAsnMetThrThrAsn	121
Db	325	ATGAACCTACGGCAGCTACATGGAGGAGAAGCACATGCCACCCCAACATGACCACGAAC	384
QY	122	GluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGln	141
Db	385	GAGCGCAGAGTTATCGTCCGACGAGATCTTACGCTTAGAGTACAGCACTGTGGCGAG	444
QY	142	TrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPheGln	161
Db	445	TGGCTGGAGTGGCGGTGAAGAATATGGCTTCCAGACGTCAACATCTTTGTTATTCCAG	504
QY	162	AsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThrPro	181
Db	505	AACATCGATGGAGGAACTGTGCAAGATGCCAAGGACGACTTCCAGAGCTCACCCTCC	564
QY	182	SerTyrAsnAlaAspIleLeuSerHisLeuHisTyrLeuArgGluArgGlyAlaThr	201
Db	565	AGCTACAAACGGCAGCATCTCTCTCATCTCCACTACCTCAGAGAGACTCTCTTCCA	624
QY	202	PheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrThrArgPro	221
Db	625	CATTG-----ACITCA-----GATGATTTGATAAGCCTTACAAAACCTTCCA	669
QY	222	-----AspLeuProTyrGluGlnAlaArgArgSerAlaTyr	233
Db	670	CGGTTATGTCATGCTAGAAACACAGATTTACCATATGAGCCCCCAGAGATCAGCCTGG	729
QY	234	ThrSerHisSerHisProThr---GlnSerLysAlaThrGlnProSerSerSerThrVal	252
Db	730	ACCGGTACGGCCACCCACGCCCTCAGTGAAGCTCTCAACCATCTCTCTCCACAGTG	789
QY	253	ProLysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSer	272
Db	790	CCCAAACTGAAGACCCAGCGTCTCTCAGTTAGATCTTATCAGATTCTTGGACCAACAGT	849
QY	273	SerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeu	292
Db	850	AGCCGCTTGCAAAATCCAGGAGTGGCCAGATCCAGCTTTGGCAGATTCTCTCTGGAGCTC	909
QY	293	LeuSerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLys	312
Db	910	CTGTGGACAGCTCCAACTCCAGCTGCATCACTGGGAAGGCCACCAACGGGGGATTCAAG	969
QY	313	MetThrAspProAspGluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMet	332
Db	970	ATGACGGATCCGACAGAGTGGCGCGCTGGGAGAGCGGAAGCAACCAACATG	1029
QY	333	AsnTyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLys	352
Db	1030	AACACGATAGCTCAGCCGCGCCTCCGTTACTTACTATGACAAGAACATCATGACCAAG	1089

Qy 353 ValHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGln 372  
Db 1090 GTCCATGGGAAGCGCTACGCTACAGTTCGACTTCACGGGATGCGCCAGGCCCTCCAG 1149  
Qy 373 ProHisProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSer 392  
Db 1150 CCCACCCCGGGAGTCACTCTGTACAGTACCCCTCAGACTCCGNCARIGGGCTCC 1209  
Qy 393 TyrHisAlaHisProGlnLysMetAsnPheValAlaProHisProAlaLeuProVal 412  
Db 1210 TATCACGCCCCACCCACAGAGATGAATTTGTGGCGCCGCCACCTCCAGGCCCTCCCGGTG 1269  
Qy 413 ThrSerSerPhePheAlaProAsnProTyrTrpAsnSerProThrGlyGlyIle 432  
Db 1270 ACATCTTCCAGTTTGTGCTGCCCAACCCACTACTGGAAATCACCACATGGGGGTATA 1329  
Qy 433 TyrProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451  
Db 1330 TACCCACACATAGCTCCGCCACAGCCATATGCTTCTCATCTGGGCACTTACTAC 1386  
RESULT 5  
ID ABX76295  
XX ABX76295 standard; DNA; 3166 BP.  
XX AC ABX76295;  
XX DT 02-APR-2003 (first entry)  
XX Lung cancer-associated polynucleotide #159.  
XX Kw Lung cancer-associated polynucleotide; gene; ds; cytostatic; emphysema;  
XX Kw antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;  
XX Kw small cell lung cancer; benign lesion; precancerous lesion; bronchitis;  
XX Kw chronic obstructive pulmonary disease; hypersensitivity pneumonitis;  
XX Kw interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.  
XX Unidentified.  
XX OS  
XX WO200286443-A2.  
XX PN  
XX 31-OCT-2002.  
XX PD  
XX PF 18-APR-2002; 2002WO-US012476.  
XX PR 18-APR-2001; 2001US-0284770P.  
XX PR 10-MAY-2001; 2001US-0290492P.  
XX PR 09-NOV-2001; 2001US-0339245P.  
XX PR 13-NOV-2001; 2001US-0350665P.  
XX PR 29-NOV-2001; 2001US-0334370P.  
XX PR 12-APR-2002; 2002US-0372246P.  
XX PA (EOSB-) EOS BIOTECHNOLOGY INC.  
XX PI Aziz N, Murray R;  
XX DR WPI; 2003-093161/08.  
XX DR P-PSDB; ABU56566.  
XX PT Detecting a lung cancer-associated transcript in a cell from a patient  
XX PT for treating lung cancer, by contacting a biological sample from the  
XX PT patient with a polynucleotide that exhibits increased or decreased  
XX PT expression in lung cancer.  
XX PS Claim 22; Page 310-311; 453pp; English.  
XX CC The invention relates to a method for detecting a lung cancer-associated  
XX CC transcript in a cell from a patient, comprising contacting a biological  
XX CC sample from the patient with a polynucleotide that selectively hybridises  
XX CC to a sequence that is at least 80 % identical to a gene that exhibits  
XX CC increased or decreased expression in lung cancer samples. Lung cancer-  
XX CC associated polynucleotides and polypeptides are used for identifying a  
XX CC compound that modulates a lung cancer-associated polypeptide, for  
XX CC inhibiting proliferation of a lung cancer-associated cell to treat lung

CC cancer in a patient and for treating a mammal having lung cancer by  
CC administering a modulatory compound identified. The methods are useful  
CC for treating lung cancer, such as small cell lung cancer, non-small cell  
CC lung cancer or other benign or precancerous lesions, e.g. atelectasis,  
CC emphysema, bronchitis, chronic obstructive pulmonary disease, fibrosis,  
CC hypersensitivity pneumonitis, interstitial pulmonary fibrosis, asthma and  
CC bronchiectasis. The genes, polynucleotides and polypeptides are useful  
CC for diagnostic purposes and as targets for screening for therapeutic  
CC compounds that modulate lung cancer, such as antibodies. Sequences  
CC ABX76124-ABX76474 represent lung cancer-associated polynucleotides of the  
CC invention  
XX  
SQ Sequence 3166 BP; 915 A; 770 C; 768 G; 713 T; 0 U; 0 Other;  
Alignment Scores:  
Pred. No.: 1.55e-176 Length: 3166  
Score: 2182.00 Matches: 407  
Percent Similarity: 91.72% Conservative: 14  
Best Local Similarity: 88.67% Mismatches: 24  
Query Match: 89.28% Indels: 14  
Df: 7 Gaps: 4  
US-09-902-772-2 (1-451) x ABX76295 (1-3166)  
Qy 2 AlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGluCys 21  
Db 281 GCAGCTCATATCAGAGAGCCCTTATCAGTTGTGAGTGAGGACCAGTCGTGTTTGGTGT 340  
Qy 22 AlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerSerGluTyr 41  
Db 341 GCCTACGGAGCCACACCTGGCTAAGACAGAGATGACCGCTCTCTCTCCAGCACTAT 400  
Qy 42 GlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnProPro 61  
Db 401 GGACAGACTTCCAGATGAGCCACGCTCCCTCAGCAGGATTGGTGCTCAACCCCA 460  
Qy 62 AlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSer 81  
Db 461 GCCAGGCTCACCATCAAAATGGAATGTAACCTAGCCAGGTGAATGGCTCAAGGAACCT 520  
Qy 82 ProAspAspCysSerValAlaLysGlyLysMetValSerSerSerSerAspAsnValGly 101  
Db 521 CCTGATGATGAGTGTGGCCAAAGGCGGAAGATGGTGGCGAGCCAGACACCGTTGG 580  
Qy 102 MetAsnTyrGlySerTyrMetGluLysHisIleProProProAsnMetThrThrAsn 121  
Db 581 ATGAATACGGCAGCTACATGGAGGAGAGACATGCCACCCCAACATGACCCAGAAC 640  
Qy 122 GluArgArgValIleValProAlaAspProThrLeuTyrSerThrAspHisValArgGln 141  
Db 641 GAGCGCAGAGTTATCGTGCACAGATCTACGCTATGGAGTACAGACCATGTGGGCAG 700  
Qy 142 TrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPheGln 161  
Db 701 TGGCTGGAGTGGGCGGTGAAGATATGGCCTTCCAGACGTCACACATCTTGTATTCCAG 760  
Qy 162 AsnIleAspGlyLysGluLeuLysMetThrLysAspAspPheGlnArgLeuThrPro 181  
Db 761 AACATCGATGGGAGGAACTGTGACAGATGACACAGGACGACTTCAGAGGCTCACCC 820  
Qy 182 SerTyrAsnAlaAspIleLeuSerHisLeuHisTyrLeuArgGlnArgGlyAlaThr 201  
Db 821 AGCTACACGCCGACATCTCTCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCA 880  
Qy 202 PheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrArgPro 221  
Db 881 CATTTG-----ACTTCA-----GATGATGTTGATAAGCCCTTCAAACTCTCCA 925  
Qy 222 -----AspLeuProTyrGluGlnAlaArgArgSerAlaTyr 233  
Db 926 CGGTTAATGCTAGTAACACAGATTTCACATATGAGCCGCCAGGAGATCAGCGCTGG 985  
Qy 234 ThrSerHisSerHisProThr-----GlnSerLysAlaThrGlnProSerSerThrVal 252

Db 986 ACAGTACAGGCGACCCAGCCAGTCCGAGTCTCAACATCTCTCCACAGTG 1045  
 Qy 253 ProlysthrGluAspGlnArgProGlnLeuAspProTyrGlnLeuGlyProThrSer 272  
 Db 1046 CCCAAACTGAAGACAGCGCTCTCAGTTAGATCTTATCAGATTCTTGACCAACAGT 1105  
 Qy 273 SerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuGluLeu 292  
 Db 1106 AGCGGCTTCAATCCAGGAGGCGCAGATCCAGCTTTGGCAGTCTCTCTGGAGCTC 1165  
 Qy 293 LeuSerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLys 312  
 Db 1166 CTGTGGACAGCTTCACTCCAGCTGATCATCCTGGGAGGACCAACCGGGAGTTCAAG 1225  
 Qy 313 MetThrAspProAspGluAlaAlaArgArgTrpGlyGluArgLysSerLysProAsnMet 332  
 Db 1226 ATGACGATCCGACGAGGTGGCCGCGCTGGGAGAGCGAGAGCAACCAACATG 1285  
 Qy 333 AsnTyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLys 352  
 Db 1286 AACTACGATAGCTACGCGCGGCTCCGTACTACTATGACAGACATCATGACCAAG 1345  
 Qy 353 ValHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGln 372  
 Db 1346 GTCCATGGAGCGCTACGCTACAGTTCCAGTCTCCACGGATCGCCAGGCGCTCCAG 1405  
 Qy 373 ProHisProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSer 392  
 Db 1406 CCCACCCCGGAGTCTCTCTGATCAAGTACCCCTCAGACCTCCCGTACATGGGCTCC 1465  
 Qy 393 TyrHisAlaHisProGlnLysMetAsnPheValAlaProHisProAlaLeuProVal 412  
 Db 1466 TATCACGCCCCACCCACAGAGATGAACCTTTGTGGCGCCGCCACCTCCAGCGCTCCCGGTG 1525  
 Qy 413 ThrSerSerSerPhePheAlaAlaProAsnProTyrTrpAsnSerProThrGlyGlyIle 432  
 Db 1526 ACAFTCTCCAGTCTTTTGTGTCGCCCAACCCATCTAGTGAATTCACCACTGGGGGTATA 1585  
 Qy 433 TyrProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451  
 Db 1586 TACCCCAACACTAGGCTCCCAACAGCCATATGCTTCTCATCTGGGCACTTACTAC 1642  
 RESULT 6  
 ADB75274  
 ID ADB75274 standard; cDNA; 3166 BP.  
 AC ADB75274;  
 XX 04-DEC-2003 (first entry)  
 DE Prostate cancer marker cDNA.  
 XX Prostate; cancer; cytostatic; gene therapy; marker; ss.  
 OS Homo sapiens.  
 XX WO2003009814-A2.  
 XX 06-FEB-2003.  
 XX 25-JUL-2002; 2002WO-US023913.  
 XX 25-JUL-2001; 2001US-0307982P.  
 XX 22-AUG-2001; 2001US-0314356P.  
 XX 25-SEP-2001; 2001US-0325020P.  
 XX 12-DEC-2001; 2001US-0341746P.  
 XX 05-MAR-2002; 2002US-0362159P.  
 XX (MILL-) MILLENNIUM PHARM INC.  
 XX Schlegel R, Monahan JE, Endege WO, Gannavarapu M, Gorbacheva B;  
 PI Hoersher S, Kamatkar S, Woney AM, Glatt K, Zhao X, Anderson D;

XX WPI; 2003-248033/24.  
 DR New nucleic acid molecule, useful for diagnosing or treating prostate cancer.  
 PT Disclosure; SEQ ID NO 98; 99pp; English.  
 PS The invention relates to newly discovered cancer markers associated with the cancerous state of prostate cells. Also disclosed is a method of assessing whether a patient is afflicted with prostate cancer. The method of the invention involves assessing whether a patient is afflicted with prostate cancer by comparing the level of expression of a marker in a patient sample and the normal level of expression of the marker in a control non-prostate cancer sample, where a significant increase in the level of expression of the marker in the patient sample and the normal level indicates that the patient is afflicted with prostate cancer.  
 CC Nucleic acids of the invention are useful for diagnosing or treating prostate cancer, and may be useful in gene therapy. Sequences given in ADB75177-ADB75631 represent marker cDNA and proteins. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence 3166 BP; 915 A; 770 C; 768 G; 713 T; 0 U; 0 Other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 1.55e-176 Length: 3166  
 Score: 2182.00 Matches: 407  
 Percent Similarity: 91.72% Conservative: 14  
 Best Local Similarity: 88.67% Mismatches: 24  
 Query Match: 99.28% Indels: 14  
 DB: Gaps: 4  
 US-09-902-772-2 (1-451) x ADB75274 (1-3166)  
 Qy 2 AlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGluCys 21  
 Db 281 CGAGCTCATATCAGAAAGCCCTTATCAGTTGTGAGTGAGGACCACTGTTGTTGAGTGT 340  
 Qy 22 AlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerSerGluTyr 41  
 Db 341 GCCTACGGAAACGCCACACTGGCTAAGACAGAGATGACGGGTCTCTCTCCACGACTAT 400  
 Qy 42 GlyGlnThrSerLysMetSerProArgValProGlnGlnAsnTrpLeuSerGlnProPro 61  
 Db 401 GGACAGACTTCCAGATGAGCCACGCGTCCCTCAGCAGGATGGCTGTCTCACACCCCA 460  
 Qy 62 AlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSer 81  
 Db 461 GCCAGGGTCAACCATCAAAATGGAATGTAACCTAGCCAGGTGAATGGCTCAAGGAATCT 520  
 Qy 82 ProAspAspCysSerValAlaLysGlyLysMetValSerSerSerSerAspAsnValGly 101  
 Db 521 CCTGATGAATGAGTGTGGCCAAAGCGGGAGAGATGGTGGGCGAGCCCGACACCGTTGGG 580  
 Qy 102 MetAsnTyrGlySerTyrMetGluLysHisIleProProAsnMetThrThrAsn 121  
 Db 581 ATGAATACCGCAGCTACATGAGGAGAGACACATGCCACCCCAACATGATGACCAAGAC 640  
 Qy 122 GluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGln 141  
 Db 641 GAGGCGCAGAGTTATCGTGCACAGATCTTACGCTATGAGGTACAGACCATGTGGGCGAG 700  
 Qy 142 TrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPheGln 161  
 Db 701 TGGCTGGAGTGGCGGTGAAGAATATGGCCCTTCAGACCTCAACATCTTGTATTATCCAG 760  
 Qy 162 AsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThrPro 181  
 Db 761 AACATCGATGGGAGGAACCTGTGCAAGATGACCAAGGACGACTTCCAGAGGCTCACCCCC 820  
 Qy 182 SerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArgGlyAlaThr 201





```

Db 341 TTGTTATTCAGAACATCGATGGGAAGAACTGTGCAAGATGACCAAGGACGACTTCCAG 400
QY 178 ArgLeuThrProSerTyrAsnAlaAspIleLeuSerHisLeuHisTyrLeuArgGlu 197
Db 401 AGGCTCACCCAGCTCAACGCGACATCTCTCTCACATCTCCACTACCTCCAGAG 460
QY 198 ArgGlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgile 217
Db 461 AGGGTGCAGCTTTTATTTTCCAAATACTTCTAGTATATCTGAAGCTACGCAAGAAATT 520
QY 218 ThrThrArgProAspLeuProTyrGluGlnAlaArgSerAlaThrThrSerHisSer 237
Db 521 ACACTAGCCAGATTTACCATATGAGCCCCCAGGAGATCAGCTCGACCGGTCCAGGC 580
QY 238 HisProThr---GlnSerIleAlaThrGlnProSerSerThrValProLysThrGlu 256
Db 581 CACCCACGCCAGCTGAAAGCTCTCAACATCTCTCCACAGTGCACAAACCTGAA 640
QY 257 AspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAla 276
Db 641 GACGAGCTCTCTAGTTAGATCTTTATCAGATTTCTTGACCAACAGTAGCCCTTGCA 700
QY 277 AsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeuLeuSerAspSer 296
Db 701 AATCCAGCAGTGGCCAGATCCAGCTTTGGCAGTTCTCTCTGGAGCTCTCTGCGACAG 760
QY 297 SerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspPro 316
Db 761 TCCAACTCCAGCTGATCACCCTGGGAAGGACCAACGCGGAGTTCAAGATGACGGATCCC 820
QY 317 AspGluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMetAsnTyrAspLys 336
Db 821 GACGAGGTGGCCGCGCTGGGAGAGCGGAAGCAACCCCAACATGAAGTACGATAAG 880
QY 337 LeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLys 356
Db 881 CTCAGCCGCGCCCTCCGTTACTACTATGACAAAGAACATCATGACCAAGGTCCATGGGAAG 940
QY 357 ArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProPro 376
Db 941 CGTAGCGCTCAAGTTGAGCTTCCACGGGATCCCGCAGGCCCTCCAGCCCCACCCCCCG 1000
QY 377 GluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHis 396
Db 1001 GAGTCATCTGTGTAAGTACCCCTCGAGACTCCCGTACATGGGCTCTATATACGCCCCAC 1060
QY 397 ProGlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSerSerSer 416
Db 1061 CCACAGAGATGAATTTGTGGCGCCCCACCTCCAGCCCTCCCGTGACATTTCCAGT 1120
QY 417 PhePheAlaAlaProAsnProTyrTrpAsnSerProThrGlyGlyIleTyrProAsnThr 436
Db 1121 TTTTGTGTCGCCCAACCCCATCTGGAAATTCACCAACTGGGGGTATATACCCCAACT 1180
QY 437 ArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
Db 1181 AGGCTCCCCACGCCATATGCTTCTCACTCTGGGCACTTACTAC 1225

RESULT 8
ID ABZ68769 standard; DNA; 1372 BP.
XX AC ABZ68769;
XX XX
DT 16-MAY-2003 (first entry)
DE Nucleotide sequence of human ERG splice variant C-1-4.
XX KW Human; splice variant; ERG; Ets; transcription factor; C-1-1; C-1-2;
KW C-1-3; C-1-4; chondrocyte; articular chondrocyte; bone disease;
KW cartilage disease; tumour; gene; ss.
XX XX
OS Homo sapiens.

```

```

XX Key
XX CDS
XX Location/Qualifiers
XX 146..1309
XX /*tag= a
XX /product= "ERG splice variant C-1-4"
XX
XX WO2003000724-A2.
XX
XX 03-JAN-2003.
XX
XX 08-MAR-2002; 2002WO-EP002505.
XX
XX 25-JUN-2001; 2001US-0300756P.
XX
XX (TIGB-) TIGENIX NV.
XX
XX Luyten F, De Bari C, Dell'accio F;
XX
XX WPI; 2003-221398/21.
XX P-PSDS; ABP97698.
XX
XX Novel splice variants of the Ets-related gene (ERG) transcription factor
XX diseases or Ets-related gene associated tumor.
XX
XX Claim 112; Fig 7; 60pp; English.
XX
XX The present sequence encodes splice variant C-1-4 of human ERG. ERG
XX belongs to the Ets family of transcription factors. The specification
XX describes C-1-1 (characterised by the absence of the 81 bp fragment, and
XX the presence of the 218 bp and 72 bp fragments of the ERG-3 cDNA
XX sequence); C-1-2 (characterised by the absence of the 81 bp and 218 bp
XX fragments); and the presence of the 72 bp fragment of the ERG-3 cDNA
XX sequence); C-1-3 (characterised by the absence of the 72 bp, 81 bp and
XX 218 bp fragments of the ERG-3 cDNA sequence); and C-1-4 (characterised by
XX the absence of the 218 bp fragment, and the presence of the 72 bp and 81
XX bp fragments of the ERG-3 cDNA sequence). The splice variant
XX polynucleotides are useful for stabilizing the phenotype of chondrocytes
XX or cells with a chondrocytes-like phenotype, for preventing de-
XX differentiation of articular chondrocytes during in vitro culture
XX expansion, for preventing or treating bone-related or cartilage-related
XX diseases or Ets-related gene associated tumour, or as marker of
XX progenitor cells of stable chondrocytes. Cells expressing the
XX polynucleotide are useful for producing or repairing cartilage tissue in
XX a mammal
XX
XX Sequence 1372 BP; 359 A; 430 C; 314 G; 269 T; 0 U; 0 Other;

```

```

Alignment Scores:
Pred. No.: 4.02e-155 Length: 1372
Score: 1926.00 Matches: 355
Percent Similarity: 89.80% Conservative: 6
Best Local Similarity: 88.31% Mismatches: 13
Query Match: 78.81% Indels: 28
DB: 8 Gaps: 2

```

US-09-902-772-2 (1-451) x ABZ68769 (1-1372)

```

QY 78 SerArgAsnSerProAspAspCysSerValAlaLysGlyLysMetValSerSerSer 97
Db 101 TCAAGGAACCTCTCTGATGAATGATGCGCCAAAGCGCGGAGATGTTGGCGCAGCCA 160
QY 98 AspAsnValGlyMetAsnTyrGlySerTyrMetGluGlyLysHisIleProProAsn 117
Db 161 GACACCGTTGGATGAAGTACGAGTACGAGTACATGAGGAGAGACATGCCACCCCAAC 220
QY 118 MetThrThrAsnGluArgArgValIleValProAlaAspProThrLeuTrpSerThrAsp 137
Db 221 ATGACCAAGAACGAGCGCAGAGTATCTGTCAGCAGATCTCTACGTATGAGTACAGAC 280
QY 138 HisValArgGlnTrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIle 157
Db 281 CATGTGGCGAGTGGCTGGAGTGGCGGTGAAGATATGSCCTTCCAGACGTCAACATC 340

```



QY	158	LeuLeuPheGlnAsnIleAspGlyIysGluLeuCysLysMetThrLysAspPheGln	177
DB	341	TTGTTATTCCAGAAACATCATGCGGAAGGAACCTGTGCAAGATGACCAAGACGACTTCCAG	400
QY	178	ArgLeuThrProSerTyrAsnAlaAaspIleLeuLeuSerHisIstleuHisTyrLeuArgGlu	197
DB	401	AGGCTCACCCCCAGCTACAAACGCCGACATCCTTCTCTCACAATCTCCACTACTCTCAGAG	460
QY	198	Arg-----	198
DB	461	ACTCTCTTCCACATTTGACTTCAGATGATGTTGTATAAAGCCTTACAAAACCTCTCCACGG	520
QY	199	-----GlyAlaThrPheIlePheProAsnThrSerValTyr	210
DB	521	TTAATGCATGCTAGAAACACAGGGGTGAGCTTTTATTTCCTCCAATCTCTCAGATAT	580
QY	211	ProGluAlaThrGlnArgIleThrThrArgProAspLeuProTyrGluGlnAlaArgArg	230
DB	581	CCTGAAGCTACGCCAAGAATTACAATAGGCCAGATTTACCATATGAGCCCCCAGGAGA	640
QY	231	SerAlaTrpThrSerHisSerHisProThr--GlnSerLysAlaThrGlnProSerSer	249
DB	641	TCAGCTCGGACCGGTCAACGCCACCCACGCCAGTCGGAAGACTGCTCAACCATCTCCT	700
QY	250	SerThrValProLysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGly	269
DB	701	TCCACAGTGCCCAAACTGAAGACCGAGCTCTCAGTTAGATCCTTATCAGATCTTGG	760
QY	270	ProThrSerSerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeu	289
DB	761	CCAACAAGTAGCCGCTTGCAAAATCCAGGCAGTGGCCAGATCCAGCTTTGGCAGTTCCCT	820
QY	290	LeuGluLeuLeuSerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGly	309
DB	821	CTGAGAGTCTCTGTCCGACAGCTCCAATCTCAGCTGCATCACTCTGGGAGGCAACCAAGGG	880
QY	310	GluPheLysMetThrAspProAspGluValAlaArgArgTrpGlyGluArgLysSerLys	329
DB	881	GAGTTCAAGATGACGGATCCCGACAGGTGGCCGGCGCTGGGGAGAGCGGAAGAGCAAA	940
QY	330	ProAsnMetAsnTyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIle	349
DB	941	CCCAACATGAACTACGATAAGCTCAGCGCGGCCCTCCGTACTACTATGATCAAGAAGATC	1000
QY	350	MetThrLysValHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGln	369
DB	1001	ATGACCAAGGTCCATGGGAAGCGCTACGCTACAAGTTCGACTTCCACGGGATCGCCAG	1060
QY	370	AlaLeuGlnProHisProProGluSerSerMetTyrLysTyrProSerAspLeuProTyr	389
DB	1061	GCCCTCCAGCCCCACCCCCCGAGTCACTCTGTACAAGTACCCCTCAGACCTCCCGTAC	1120
QY	390	MetSerSerTyrHisAlaHisProGlnLysMetAsnPheValAlaProHisProProAla	409
DB	1121	ATGGGCTCCTATCAGCCCCACCCACAGAGATGAACTTTGTGGCGCCCCACCTCCAGCC	1180
QY	410	LeuProValThrSerSerSerPhePheAlaAlaProAsnProTyrTrpAsnSerProThr	429
DB	1181	CTCCCCGTGACATCTTCAGTTTTTTTGTCTGCCCAACCCCATCTCGAAATTCACCAACT	1240
QY	430	GlyGlyIleTyrProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThr	449
DB	1241	GGGGGTATATACCCCAACACTAGGCTCCCCACGACCATATGCCCTTCTCATCTGGGCAC	1300
QY	450	TyrTyr-----	451
DB	1301	TTACTAC 1306	

RESULT 9  
 ABZ68768  
 standard; DNA; 1219 BP.  
 XX

AC	AB268768;
XX	
DT	16-MAY-2003 (first entry)
XX	
DE	Nucleotide sequence of human ERG splice variant C-1-3.
XX	
KW	Human; splice variant; ERG; Ets; transcription factor; C-1-1; C-1-2;
XX	
KW	C-1-3; C-1-4; chondrocyte; articular chondrocyte; bone disease;
XX	
KW	cartilage disease; tumour; gene; ss.
XX	
OS	Homo sapiens.
XX	
Key	Location/Qualifiers
FF	146..1156
CDS	/*tag= a
FT	/product= "ERG splice variant C-1-3"
FT	
FT	
XX	
XX	WO2003000724-A2.
PN	
XX	
XX	03-JAN-2003.
XX	
XX	08-MAR-2002; 2002WO-EP002605.
XX	
PF	
XX	
PR	25-JUN-2001; 2001US-0300756P.
XX	
XX	(TIGEN) TIGENIX NV.
PA	
XX	
PI	Luyten F, De Bari C, Dell'accio F;
XX	
XX	WPI: 2003-2241398/21.
DR	P-PGDB; ABP97697.
DR	
XX	
XX	Novel splice variants of the Ets-related gene (ERG) transcription factor
XX	
PT	, useful for preventing or treating bone-related or cartilage-related
PT	diseases or Ets-related gene associated tumor.
PT	
XX	
XX	Claim 11; Fig 6; 60pp; English.
PS	
XX	
XX	The present sequence encodes splice variant C-1-3 of human ERG. ERG
CC	
CC	belongs to the Ets family of transcription factors. The specification
CC	describes C-1-1 (characterised by the absence of the 81 bp fragment, and
CC	the presence of the 218 bp and 72 bp fragments of the ERG-3 cDNA
CC	sequence); C-1-2 (characterised by the absence of the 81 bp and 218 bp
CC	fragments, and the presence of the 72 bp fragment of the ERG-3 cDNA
CC	sequence); C-1-3 (characterised by the absence of the 72 bp, 81 bp and
CC	218 bp fragments of the ERG-3 cDNA sequence); and C-1-4 (characterised by
CC	the absence of the 218 bp fragment, and the presence of the 72 bp and 81
CC	bp fragments of the ERG-3 cDNA sequence). The splice variant of chondrocytes
CC	polynucleotides are useful for stabilising the phenotype of chondrocytes
CC	or cells with a chondrocytes-like phenotype, for preventing de-
CC	differentiation of articular chondrocytes during in vitro culture
CC	expansion, for preventing or treating bone-related or cartilage-related
CC	diseases or Ets-related gene associated tumour, or as marker of
CC	progenitor cells of stable chondrocytes. Cells expressing the
CC	polynucleotide are useful for producing or repairing cartilage tissue in
CC	a mammal
XX	
SQ	Sequence 1219 BP; 312 A; 393 C; 288 G; 226 T; 0 U; 0 Other;
Alignment Scores:	
Pred. No.:	4,74e-145 Length: 1219
Score:	1807.50 Matches: 332
Percent Similarity:	90.40% Conservative: 7
Best Local Similarity:	88.53% Mismatches: 11
Query Match:	73.96% Indels: 25
DB:	Gaps: 2
XX	
US-09-902-772-2 (1-451) x AB268768 (1-1219)	
Oy	78 SerArgAsnSerProAspAspCysSerValAlaLysGlyLysMetValSerSerSer 97
Db	101 TCAAGGAATCTCTCTGATGAAATGCTAGTGTGTCCTCAAGCGGGAAGATGCGGCAGCCCA 160

Qy 98 AspAsnValGlyMetAsnTyrGlySerTyrMetGluGluHisIleProProProAsn 117  
 Db 161 GACACCTTGGATGAATACGCGCAGCTACATGAGGAGGAGCAGCATGGCCCCCAAC 220  
 Qy 118 MetThrThrAsnGluArgArgValIleValProAlaAspProThrLeuTyrSerThrAsp 137  
 Db 221 ATGACACGAGACGAGCGAGATTTATGTCGCCAGCAGATCTTACGCTATGGATGACAGAC 280  
 Qy 138 HisValArgGlnTrpLeuGluTrpAlaValIleGluTyrGlyLeuProAspValAspIle 157  
 Db 281 CATGTGGCGAGTGGCTGGAGTGGCGGTGAAGAATATGGCTTCCAGACGCTCAACATC 340  
 Qy 158 LeuLeuPheGlnAsnIleAspGlyIleGluLeuCysIleMetThrIleValAspAspPheGln 177  
 Db 341 TTGTATTCAGAACATCGATGGGAGAACTGTGCAAGATGACCAAGGACGACTTCCAG 400  
 Qy 178 ArgLeuThrProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGlu 197  
 Db 401 AGGCTCACCCCGCAGTACACCGCGACATCTTCTCTCACATCTCCACTACCTCAGAGAG 460  
 Qy 198 ArgGlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIle 217  
 Db 460 ----- 460  
 Qy 218 ThrThrArgProAspLeuProTyrGluGlnAlaArgSerAlaTrpThrSerHisSer 237  
 Db 461 -----AATTACCATATGAGCCCCCGCAGGAGATCAGCTGGACCGGTCACGGC 508  
 Qy 238 HisProThr-----GlnSerLysAlaThrGlnProSerSerSerThrValProLysThrGlu 256  
 Db 509 CACCCCGCGCCGAGTGGAAAGCTGCTCAACCATCTCTTCCAGAGTCCCAAACTGAA 568  
 Qy 257 AspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAla 276  
 Db 569 GACACAGCTCTCAGTTAGATCTTATCAGATTTCTTGACCAACAGTACCGCGCTTGA 628  
 Qy 277 AsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeuLeuSerAspSer 296  
 Db 629 AATCCAGGACGTGGCCAGATCCAGCTTGGCAGTCTCTCTGGAGCTCCTGTGGACAGC 688  
 Qy 297 SerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspPro 316  
 Db 689 TCCAACTCCAGCTGCATCACCCTGGAGGACCAACCGGGAGTTCAAGATGACGGATCCC 748  
 Qy 317 AspGluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMetAsnTyrAspLys 336  
 Db 749 GACGAGGTGGCGCGCTGGGAGAGCGGAGAGCAACCCACATGAACTACGATAAG 808  
 Qy 337 LeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLys 356  
 Db 809 CTCAGCCGCGCCCTCGGTACTATGATGACAGAACATCATGACCAAGGTCCATGGGAAG 868  
 Qy 357 ArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProPro 376  
 Db 869 CGTACGCTCAAGTTGCACTTCCAGGGATGCGCCAGCGCTTCCAGCCCGCCACCCCGC 928  
 Qy 377 GluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHis 396  
 Db 929 GAGTCACTCTGTACAGTACCCCTCAGACCTCCCGTACATGGGCTCTCTATCAGCGCCAC 988  
 Qy 397 ProGlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSerSerSer 416  
 Db 989 CCACAGAAAGTGAATTTGTGGCGCCCGCCCGCTCCAGCCCTCCCGGTGACATCTTCCAGT 1048  
 Qy 417 PhePheAlaAlaProAsnProTyrTrpAsnSerProThrGlyIleTyrProAsnThr 436  
 Db 1049 TTTTGTGTCGCCCAACCCCATCTGGAATTCACCACTGGGGGTATATACCCCAACACT 1108  
 Qy 437 ArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451  
 Db 1109 AGGCTCCCGCACGACCATATGCTTCTCATCTGGGCATTTACTAC 1153

RESULT 10

ADA02692  
 ID ADA02692 standard; cDNA; 1359 BP.  
 XX AC ADA02692;  
 XX DT 06-NOV-2003 (first entry)  
 XX DE Human FLI1 carcinoma associated coding sequence, SEQ ID NO:1210.  
 XX KW Human; carcinoma associated; oncogene; carcinoma; cancer; breast;  
 KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;  
 KW gene; ss.  
 XX OS Homo sapiens.  
 XX PN WO2003057146-A2.  
 XX PD 17-JUL-2003.  
 XX PF 26-DEC-2002; 2002WO-US041414.  
 XX PR 26-DEC-2001; 2001US-00035832.  
 XX (SAGR-) SAGRES DISCOVERY.  
 XX PI Morris DW;  
 XX DR WPI; 2003-587068/55.  
 XX PT New recombinant nucleic acid encoding carcinoma associated protein,  
 XX useful for preparing compositions for treating carcinomas.  
 XX PS Claim 1; SEQ ID NO 1210; 245pp; English.  
 XX CC The invention relates to recombinant carcinoma associated (CA) nucleic  
 CC acid sequences from mouse and human (ADA01482-ADA03094), and to  
 CC recombinant carcinoma associated proteins (CAP) encoded by them. The  
 CC invention also encompasses expression vectors and host cells comprising a  
 CC CA nucleic acid, a polypeptide (especially an antibody) that specifically  
 CC binds to the protein, and a biochip comprising CA nucleic acid or  
 CC fragments thereof. The sequences of the invention were identified using  
 CC oncogenic retroviruses, which insert into the genome of the host organism  
 CC at random. Many of these do not carry transduced host oncogenes or  
 CC pathogenic trans-acting viral genes, meaning that cancer incidence is a  
 CC direct consequence of the effects of proviral integration into host  
 CC protooncogenes. The CA nucleic acid sequences can be used to diagnose  
 CC carcinoma (especially breast cancer, prostate cancer, lymphoma or  
 CC leukaemia) or a propensity to carcinoma by determination of the sequence  
 CC of a CA gene, or by determination of CA gene expression in particular  
 CC tissues. CA nucleic acids, proteins and antibodies are also useful as  
 CC therapeutic agents and in screening and evaluating drug candidates. The  
 CC present sequence represents a specifically claimed human CA nucleic acid  
 CC sequence of the invention. Note: The complete sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 1359 BP; 359 A; 422 C; 333 G; 245 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 2,52e-129 Length: 1359  
 Score: 1624.50 Matches: 306  
 Percent Similarity: 79.43% Conservative: 57  
 Best Local Similarity: 66.96% Mismatches: 83  
 Query Match: 66.47% Indels: 11  
 DB: Gaps: 7  
 US-09-902-772-2 (1-451) x ADA02692 (1-1359)  
 Qy 1 MetAlaSerThrIleLysGluAlaLeuSerValSerGluAspGlnSerLeuPheGlu 20  
 Db 1 ATGACGGGACTATTAAAGAGGCTCTGTCGGTGGAGGACGACGACGCTCTCTTTGAC 60

QY 21 CysAlaTyrGly---SerProHisLeuAlaIysThrGluMetThrAlaSerSerSer 39  
 Db 61 TCAGCGTACGAGCGGCGAGCCCATCTCCCAAGCCGACATGACTGCTCGGGAGTCCT 120  
 QY 40 GluTyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGln 59  
 Db 121 GACTACGGGAGCCCGACAGATCAACCCCTCCACACAGAGAGTGTATCAATCAG 180  
 QY 60 ProProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArg 79  
 Db 181 ---CCAGTGGGGTCAACGTCAGCGGAGTAT-----GACCACATGAATGGATCCAGG 231  
 QY 80 AsnSerProAspCysSerValAlaIysGlyLysMetValSerSerSerAspAsn 99  
 Db 232 GAGTCTCGGTGACTCGAGCTTAGCAATGACGAGCTGTGGCGAGCGGAGTCC 291  
 QY 100 ValGlyMetAsnTyrGlySerTyrMetGluGluLysHis---IleProProAsnMet 118  
 Db 292 AACCCCATGAACACAGCATATATGAGCAGAGAAGATGGCCCCCTCTCCCAACATG 351  
 QY 119 ThrThrAsnGluArgArgValIleValProAlaAspProThrLeuTyrSerThrAspHis 138  
 Db 352 ACCACCAAGAGAGAGAGTATCGTCCCGCAGACCCACACATGTGACACAGGAGCAT 411  
 QY 139 ValArgGlnTyrLeuGluTyrAlaValLysGluTyrGlyLeuProAspValAspIleLeu 158  
 Db 412 GTGAGGCAATGGCTGGAGTGGCCATAAAGAGTACAGCTTGATGGAGATCGACATCC 471  
 QY 159 LeuPheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArg 178  
 Db 472 TTTTTCAGAACATGGATGGCAGGAACGTGTAAATGAACAGGAGGACTTCTCCCGC 531  
 QY 179 LeuThrProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArg 198  
 Db 532 GCCACCCCTCTACACACGGAAGTGTGTGTACACCTCAGTTACCTCAGGAA-- 588  
 QY 199 GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThr 218  
 Db 589 ---AGTTCACGTGCTGCTATTAACAACTCCACACCGCAACCACTCTCAGCATTTAGT 645  
 QY 219 ThrArgProAspLeuProTyrGluGlnAlaArgArgSerAlaTyrThrSerHisSerHis 238  
 Db 646 GTCAAGAGACCCCTCTATGATCAGTCAGTACAGAGAGGAGCTTGGGCAATACATGAT 705  
 QY 239 ProThrGlnSerLysAlaThrGln---ProSerSerSerThrValProLysThrGluAsp 257  
 Db 706 TCTGCTCTCAACAAAGTCTCCCTTGGAGGGGCGACAAACGATCAGTAAGAATACAGAG 765  
 QY 258 GlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerArgLeuAlaAsn 277  
 Db 766 CACGGCCCCAGCCAGATCCGTATCAGATCCTGGGCGGACACAGCATGCTAGCCAC 825  
 QY 278 ProGlySerGlyGlnIleGlnLeuTyrPheLeuLeuGluLeuLeuSerAspSerSer 297  
 Db 826 CTTGAAGCGGCGAGATCCAGCTGTGGCAATCTCTCTGGAGCTGCTCCGACAGGGCC 885  
 QY 298 AsnSerAsnCysIleThrTyrGluGlyThrAsnGlyGluPheLysMetThrAspProAsp 317  
 Db 886 AACGCCACTGTATCACTTGGAGGGGACCAACGGGGAGTTCAAAATGACGACCCCGAT 945  
 QY 318 GluValAlaArgArgTyrGlyLysSerLysProAsnMetAsnTyrAspLysLeu 337  
 Db 946 GAGGTGGCCAGCGCTGGGGGAGCGGAAAGCAAGCCCAACATGAATTCAGACAGCTG 1005  
 QY 338 SerArgAlaLeuArgTyrTyrThrLysAsnIleMetThrLysValHisGlyLysArg 357  
 Db 1006 AGCCGGGCGCTCGTTATTTACTATATAAAACATTTATGACCAAGTGCAGCGCAAGA 1065  
 QY 358 TyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGlu 377  
 Db 1066 TATGTTTACAAATTTGATCTCCAGCGCATTCGCCAGGCTCTGCAGCCACATCCGACGAG 1125  
 QY 378 SerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisPro 397

Db 1126 TCGTCCATGTACAAGTACCCCTTCGACATCTCTCATGCTCTTCTACCATGCCACAG 1185  
 QY 398 GlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSerSerSerPhe 417  
 Db 1186 CAGAAGGTGAACCTTGTCCCTCCCATCCATCTCCATGCTGTCTCTCTCCAGCTTC 1245  
 QY 418 PheAlaAlaProAsnProTyrTyrAsnSerProThrGlyGlyIleTyrProAsn----- 435  
 Db 1246 TTTGAGCGGCATCAAAATCTGGACCTCCCGGGAATCTACCCACCCCAAC 1305  
 QY 436 ---ThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451  
 Db 1306 GTCCCGCCCATCTCTAACCCACGCTGCTTCACTTAGCAGCTACTAC 1356

## RESULT 11

ADB72430  
 ID ADB72430 standard; cDNA; 1359 BP.

XX ADB72430;  
 AC

XX 04-DEC-2003 (first entry)

DE Human FLII cDNA.

XX human; ss; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;  
 KW cancer; neoplasm; adenocarcinoma; sarcoma; gene.

OS Homo sapiens.

XX WO2003008583-A2.

XX 30-JAN-2003.

XX 26-DEC-2001; 2001WO-US051291.

XX 02-MAR-2001; 2001US-00798586.

XX 23-OCT-2001; 2001US-0004113.

XX 08-NOV-2001; 2001US-00052482.

XX 30-NOV-2001; 2001US-00997722.

XX 20-DEC-2001; 2001US-00034650.

PA (SAGR-) SAGRES DISCOVERY.

XX Morris DW, Engelhard EK;

XX WPI; 2003-239337/23.

XX New recombinant nucleic acid, useful for treating carcinomas, lymphomas,  
 PT cancers, neoplasm, adenocarcinoma, or sarcomas.

XX Claim 1; SEQ ID NO 258; 2304pp; English.

XX The invention relates to a novel recombinant nucleic acid comprising a  
 CC nucleotide sequence selected from any of the 660 sequences fully defined  
 CC in the specification. A polynucleotide of the invention has cytostatic  
 CC activity, and may have a use in gene therapy, or in a vaccine. The  
 CC recombinant nucleic acids and polypeptides are useful for treating  
 CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and  
 CC sarcomas. The present sequence represents a human cDNA of the invention.

XX Sequence 1359 BP; 359 A; 422 C; 333 G; 245 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.:	2,52e-129	Length:	1359
Score:	1624.50	Matches:	306
Percent Similarity:	79.43%	Conservative:	57
Best Local Similarity:	66.96%	Mismatches:	83
Query Match:	9	Indels:	11
DB:		Gaps:	7

US-09-902-772-2 (1-451) x ADB72430 (1-1359)

Qy 1 MetAlaSerThrIleLysGluAlaLeuSerValIserGluAspGlnSerLeuPheGlu 20  
 Db 1 ATGACGGGACTATTAAGGAGGCTGTGCGTGGTGGAGCAGCAGCAGTCCCTCTTTGAC 60  
 Qy 21 CysAlaTyrGly---SerProHisLeuAlaLysThrGluMetThrAlaSerSerSer 39  
 Db 61 TCAGCGTACGAGCGGAGCGCCATCTCCCAAGCGCGACATGACTCCCTCGGGAGTCT 120  
 Qy 40 GlnTyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGln 59  
 Db 121 GACTACGGGAGCCCGACAGATCAACCCCTCCACACAGCAGCAGGAGTGATCAATCAG 180  
 Qy 60 ProProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArg 79  
 Db 181 ---CCAGTGGGTCAACGTCAGCGGGAGTAT-----GACCACATGAATGATCCAG 231  
 Qy 80 AsnSerProAspAspCysSerValAlaLysGlyGlyLysMetValSerSerSerAsp 99  
 Db 232 GAGTCTCCGTGAGCTGACGCTTAGCNAATGACGAGAGCTGGTGGCGGAGCGAGTCC 291  
 Qy 100 ValGlyMetAsnTyrGlySerTyrMetGluGluLysHis---IleProProAsnMet 118  
 Db 292 AACCCCATGAATACAAACAGCTATATGGACGAGAAGATGGCCCTCTCTCCCAACATG 351  
 Qy 119 ThrThrAsnGluArgValIleValProAlaAspProThrLeuTrpSerThrAspHis 138  
 Db 352 ACCACCAACAGAGGAGAGTCACTGCTCCCGCAGACCCACACTGTGGACAGGAGCAT 411  
 Qy 139 ValArgGlnTrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeu 158  
 Db 412 GTGAGGCAATGGCTGGAGTGGCCCATAAAGAGTACAGCTTGATGGAGTCGACATCC 471  
 Qy 159 LeuPheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArg 178  
 Db 472 TTTTTCAGAACATGATGGCAAGAACTGTGTAAATGAACAAGGAGGACTTCTCCGC 531  
 Qy 179 LeuThrProSerTyrAsnAlaAspIleLeuSerHisLysLysLeuArgGluArg 198  
 Db 532 GCACACACCTCTACAAACGAGGAGTGTGTTGTACACCTGATCTTACCTCAGGNA--- 588  
 Qy 199 GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThr 218  
 Db 589 ---AGTTCACTGCTGCTATAATACAACTCCACACCGACCAATCTCCACGATGAGT 645  
 Qy 219 ThrArgProAspLeuProTyrGluGlnAlaArgSerAlaTrpThrSerHisSerHis 238  
 Db 646 GTCAAGAAGACCCCTTCTTAAGTACTGATCAGAAAGAGAGCTTGGGGCAATACATGAAT 705  
 Qy 239 ProThrGlnSerLysAlaThrGln---ProSerSerSerThrValProLysThrGluAsp 257  
 Db 706 TCTGGCTCAACAAAAGTCTCCCTTGGAGGGGCACAAACGATCAGTAAGAATACAGAG 765  
 Qy 258 GlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsn 277  
 Db 766 CAACGGCCCCAGCAGATCCGTATCAGATCTCGGGCCCGACACGAGTCCCTAGGCAAC 825  
 Qy 278 ProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuLeuSerAspSerSer 297  
 Db 826 CCTGGAAGCGGGCAGATCCAGCTGTGGCAATCTCTCGAGGTGCTCTCCGACAGCGCC 885  
 Qy 298 AsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspProAsp 317  
 Db 886 AACGCCAGCTGTATCACTCGGGAGGGGACCAACGGGGAGTTCAAAATGACGGACCCCGAT 945  
 Qy 318 GluValAlaArgTrpGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeu 337  
 Db 946 GAGGTGCCAGGCGCTGGGGCGAGCGGAAAGCAAGCCCAACATGATTAACGACAGCTG 1005  
 Qy 338 SerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArg 357  
 Db 1006 AGCCGGGCCCTCGCTATTACTATGATAAAACATTATGACCAAGTGCACGGCAAAAGA 1065  
 Qy 358 TyrAlaTyrLysPheAspPheHisGlyTleAlaGlnAlaLeuGlnProHisProGlu 377

Db 1066 TATGTTACAAATTTGACTTCCACGGCATTCGCCAGGCTCTGCAGCCACATCCGACCGAG 1125  
 Qy 378 SerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisPro 397  
 Db 1126 TCGTCCATGTACAAGTACCTTCTGACATCTCTACATGCTTCTTACATGCCACCGAG 1185  
 Qy 398 GlnLysMetAsnPheValAlaProHisProAlaLeuProValThrSerSerSerPhe 417  
 Db 1186 CAGAAGGTGAACCTTTGCTCCCTCCCATCCATCTCCATGCTGTCACTTCTCCAGCTTC 1245  
 Qy 418 PheAlaAlaProAsnProTyrTrpAsnSerProThrGlyGlyIleTyrProAsn----- 435  
 Db 1246 TTTGAGCGCGCATCACATCTGCACTTCCCTCCCGGAGGAACTACCCCAACCCCAAC 1305  
 Qy 436 ---ThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451  
 Db 1306 GTCCCCCGCCATCTAACACCCAGCTGCTTACACATAGGCGAGCTACTAC 1356  
 RESULT 12  
 AAQ50644  
 ID AAQ50644 standard; cDNA; 2938 BP.  
 XX AC AAQ50644;  
 XX DT 25-MAR-2003 (revised)  
 XX DT 26-MAY-1994 (first entry)  
 XX DE Human Hum-Flt-1 gene clone BM025.  
 XX KW chromosomal translocation; chimeric; chimeraic; chimaeric; Ewing sarcoma; Ews gene;  
 XX KW malignant melanoma; hum-flt-1;  
 XX KW primitive peripheral neuroectodermal tumour; human chromosome 11;  
 XX KW human chromosome 22; ss.  
 XX OS Homo sapiens.  
 XX FH Key  
 XX FT CDS  
 XX FT Location/Qualifiers  
 XX FT 143..1501  
 XX FT /tag= a  
 XX FT polyA\_signal  
 XX FT 2908..2913  
 XX FT /tag= b  
 XX PN W09323549-A2.  
 XX PD 25-NOV-1993.  
 XX PF 19-MAY-1993; 93WO-FR000494.  
 XX PR 20-MAY-1992; 92FR-00006123.  
 XX XX (CNRS ) CNRS CENT NAT RECH SCI.  
 XX PI Aurias A, Delattre O, Desmaze C, Melot T, Peter M, Plougastel B;  
 XX PI Thomas G, Zucman J;  
 XX DR P-PSDB; AAR44556.  
 XX DR WPI; 1993-386580/48.  
 XX XX New nucleic acid of EWS gene and its hybrid(s) - contg. gene sequence  
 XX PT involved in chromosomal translocation, also derived mRNA, probes, fusion  
 XX PT proteins etc., for diagnosis and treatment of Ewing sarcoma and melanoma.  
 XX PS Disclosure; Fig 7; 123pp; French.  
 XX CC The probe 11R1 was used to screen a human marrow cDNA library (Clontech  
 XX CC cat.# HL1058). The clone EM025 was identified and sequenced. It  
 XX CC represents the entire coding region together with 5'- and 3'-UTRs of the  
 XX CC Hum-Flt-1 gene. (Updated on 25-MAR-2003 to correct PN field.)  
 XX XX Sequence 2938 BP; 847 A; 692 C; 675 G; 724 T; 0 U; 0 Other;

## Alignment Scores:

```

Pred. No.:      7,22e-129      Length:      2938
Score:          1624.50      Matches:      306
Percent Similarity: 79.43%      Conservative: 57
Best Local Similarity: 66.96%      Mismatches: 83
Query Match:      66.47%      Indels:      11
DB:              2          Gaps:       7

US-09-902-772-2 (1-451) x AAQ50644 (1-2938)

Qy      1 MetAlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20
Db      143 ATGACGGGACTATTAGAGGCTCTGTCGGTGTGAGCAGCAGCAGCAGCCTCTTGTGAC 202
Qy      21 CysAlaTyrGly---SerProHisLeuAlaLysThrGluMetThrAlaSerSerSer 39
Db      203 TCAGCGTACGAGCGGCGAGCCATCTCCCAAGCGCCGACATGACTCCCTCGGGAGTCT 262
Qy      40 GluTyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGln 59
Db      263 GACTACGGGAGCGCCCAAGATCAACCCCTCCCAAGCAGCAGGAGTGATCAATCAG 322
Qy      60 ProProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArg 79
Db      323 ---CCAGTGGGTCAACGTCAGCGGAGTAT-----GACCACATGATGATCCAGG 373
Qy      80 AsnSerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsn 99
Db      374 GAGTCTCCGTTGACTCAGCTAGCAATGCAAGCAAGCTGTGGCGGAGCGAGTCC 433
Qy      100 ValGlyMetAsnTyrGlySerTyrMetGluGluLysHis---IleProProAsnMet 118
Db      434 AACCCTGATGACTACACACTATATGGACGAGAGATGGCCCCCTCTCTCCCAATG 493
Qy      119 ThrThrAsnGluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHis 138
Db      494 ACCACCAACGAGAGAGTCACTGCTCCCGCAGAGCCACACTGTGGACACAGGAGCAT 553
Qy      139 ValArgGlnTrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeu 158
Db      554 GTGAGGCAATGCTGGAGTGGGCCATAGAGATATAGCTTGTGAGATCGACATCC 613
Qy      159 LeuPheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArg 178
Db      614 TTTTTCAGAACATGATGCAAGAACTGTGTAATAATGAACAAGAGGAGCTTCTCCGCG 673
Qy      179 LeuThrProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArg 198
Db      674 GCCACCACTCTACACACGGAAGTCTGTGTACACCTCAGTTACCTCAGGAA---730
Qy      199 GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThr 218
Db      731 --AGTTCACCTGTGGCTATATACAACTCCACACCGACCAATCCTCAGGATGAGT 787
Qy      219 ThrArgProAspLeuProTyrGluGlnAlaArgSerAlaThrThrSerHisSerHis 238
Db      788 GTCAAGAGAGACCTCTTATGACTCAGTCAGAGAGAGCATGTGGGCAATAACATGAT 847
Qy      239 ProThrGlnSerLysAlaThrGln---ProSerSerThrValProLysThrGluAsp 257
Db      848 TCTGGCTCAACAAAGTCTCCCTTGGAGGGGCACAAACGATCAGTAAGATACAGAG 907
Qy      258 GlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsn 277
Db      908 CAACGGCCCCAGCAGATCCGTATCAGATCTCGGGCCCGACGACGATCGCTAGCAAC 967
Qy      278 ProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuLeuLeuLeuSerSerSer 297
Db      968 CCTGAGAGCGGAGATCCAGCTGTGGCAATTCCTCTGAGTGTCTCTCCGACAGCGCC 1027
Qy      298 AsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspProAsp 317
Db      1028 AACGCCAGCTGTATCAGCTGGAGGGGACCAACAGGGGAGTTCAAAATGACGGACCCCGAT 1087

```

```

Qy      318 GluValAlaArgArgTrrGlyLysSerLysProAsnMetAsnTyrAspLysLeu 337
Db      1088 GAGGTGGCCAGCGCTGGGGGAGCGGAAAGCAAGCCCAACATGATTAACGACAGCTG 1147
Qy      338 SerArgAlaLeuArgTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArg 357
Db      1148 AGCCGGGCCCTCCGTTATTACTATGATAAAACATTAAGCAAAAGTGCAGCGCAAGA 1207
Qy      358 TyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGlu 377
Db      1208 TATGCTTACAAATTGACTTCCAGCGCATTCGCCAGGCTCTGCAGCCACATCCGACCGAG 1267
Qy      378 SerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisPro 397
Db      1268 TCGTCCATGTACAAGTACCTTCTGACATCTCTACATGCTCTCTACCATGCCCCACAG 1327
Qy      398 GlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSerSerSerPhe 417
Db      1328 CAGAAGTGAATTTGTCCTCCCATCCATCTCCATGCTGTCATCTCTCTCCAGCTTC 1387
Qy      418 PheAlaAlaProAsnProTyrTyrAsnSerProThrGlyGlyIleTyrProAsn----- 435
Db      1388 TTTGGAGCGGCATCACAATACTGGACCTCCGCCAGGGGGATCTACCCCAACCCCAAC 1447
Qy      436 ---ThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
Db      1448 GTCCCCCGCCATCTTAACACCCAGTGCCTTCACACTTAGCAGCTACTAC 1498

RESULT 13
ABV94304
ID ABV94304 standard; cDNA; 2957 BP.
AC ABV94304;
XX
DT 08-JAN-2003 (first entry)
XX
DE Breast carcinoma related nucleotide sequence SEQ ID NO:295.
KW Human; breast carcinoma; cancer; tumour; cytostatic; anti-tumour; gene;
XX
OS Homo sapiens.
XX
PN WO200246467-A2.
XX
PD 13-JUN-2002.
XX
PF 07-DEC-2001; 2001WO-IB002811.
XX
PR 08-DEC-2000; 2000US-0254090P.
XX
PR 07-DEC-2001; 2001US-00007926.
XX
PA (IPSO-) IPSOGEN.
XX
PI Bertucci F, Houlgatte R, Birnbaum D, Nguyen C, Viens P, Fert V;
XX
DR WPI; 2002-619023/66.
XX
PT Novel polynucleotide library useful in molecular characterization of a
PT carcinoma, comprising a pool of polynucleotide sequences or its
PT subsequences which are either underexpressed or overexpressed in tumor
PT cells.
XX
PS Claim 1; Page 300-301; 401pp; English.
XX
CC The present invention describes a polynucleotide library (I) useful in
CC the molecular characterisation of a carcinoma, comprising a pool of
CC polynucleotides or its subsequences which are either underexpressed or
CC overexpressed in tumour cells, and correspond to any of the
CC polynucleotide sequences chosen from the 488 sequences given in ABV9401.0
CC to ABV94477. Also described: (i) a polynucleotide array (II) useful for
CC the prognosis or diagnostic of tumour, comprising (I); and (2) detecting
CC

```

(M1) differentially expressed polynucleotide sequences which are correlated with a cancer, involves obtaining a polynucleotide sample from a patient, and reacting the polynucleotide sample obtained with a probe immobilised on a solid support, where the probe comprises any combination of the polynucleotide sequences of (I) or its expression products encoded by polynucleotide sequences of (I), and detecting the reaction product. (I) have cytostatic activities and can be used as anti-tumour agents. (I) is useful in molecular characterisation of a carcinoma. (I) and (II) are useful for the prognosis or diagnosis of tumour, in differentiating a normal cell from a cancer cell, detecting a hormone sensitive tumour cell, differentiating a tumour with lymph nodes from a tumour without lymph nodes, differentiating antracycline-sensitive tumours from antracycline-insensitive tumours, and classifying good and poor prognosis primary breast tumours. (I) is useful for large-scale molecular characterisation of breast cancer that help in prediction, prognosis and cancer treatment, and for detecting differentially expressed genes that correlated with a cancer

XX Sequence 2957 BP; 855 A; 698 C; 680 G; 724 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 7,28e-129 Length: 2957  
Score: 1624.50 Matches: 306  
Percent Similarity: 79.43% Conservative: 57  
Best Local Similarity: 66.96% Mismatches: 83  
Query Match: 66.47% Indels: 11  
DB: 6 Gaps: 7

US-09-902-772-2 (1-451) x ABV94304 (1-2957)

Qy 1 MetAlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20  
Db 173 ATGACCGGACTATTAGAGGGCTCTGCGTGGTGGAGCAGCAGCCAGTCTTTGAC 232  
Qy 21 CysAlaTyrGly---SerProHisLeuAlaLysThrGluMetThrLalaSerSerSer 39  
Db 233 TCAGCGTACGAGGCGGAGCCCATCTCCCAAGCGCGACATGACTGCTCGGGAGTCTCT 292  
Qy 40 GluTyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGln 59  
Db 293 GACTACCGGCGAGCCCAAGATCAACCCCTCCACAGCAGGAGGATGATCAATCAG 352  
Qy 60 ProProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArg 79  
Db 353 ---CCAGTGGGTCAACGTCAAGCGGAGTAT-----GACCACATGAATGATCCAGG 403  
Qy 80 AsnSerProAspAspCysSerValAlaLysGlyGlyLysMetValSerSerSerAspAsn 99  
Db 404 GAGTCTCCGTGGACTCAGCGTTAGCAATGAGCAAGCTGTGGCGGAGCGAGTCC 463  
Qy 100 ValGlyMetAsnTyrGlySerTyrMetGluGluLysHis---IleProProAsnMet 118  
Db 464 AACCCCATGAATACACACTATATGGACGAGAGAAATGGCCCCCTCTCTCCCAACATG 523  
Qy 119 ThrThrAsnGluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHis 138  
Db 524 ACCACCAACAGAGGAGGAGTATCGTCCCGCAGACCCCACTGTGGACAGGAGCAT 583  
Qy 139 ValArgGlnTrpLeuGluTTPAlaValLysGluTyrGlyLeuProAspValAspIleLeu 158  
Db 584 GTGAGCAATGGTGGTGGTGGCCATAAAGAGTACAGTTGTAGTGGATCGACATCC 643  
Qy 159 LeuPheGlnAsnLeuAspGlyLysGluLeuCysGlyMetThrLysAspAspPheGlnArg 178  
Db 644 TTTTTCAGAACATGATGTCGCAAGAACTGTGTAAATGAACAGGAGGACTTCTCTCCG 703  
Qy 179 LeuThrProSerTyrAnAlaAspIleLeuLeuSerHisHisTyrLeuArgGluArg 198  
Db 704 GCACACACCTCTACACACGAGGAGTCTGTGTACACCTCAGTACCTTCAGGAA--- 760  
Qy 199 GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThr 218  
Db 761 ---AGTTCACTGCTGGCCTATATACAACTCCACACCGACCAATCCTCACGATGAGT 817

Qy 219 ThrArgProAspLeuProTyrCluclnAlaArgSerAlaTTPThrSerHisSerHis 238  
Db 818 GTCAAGAGAGCCCTCTTATGACTCACTCAGAGAGAGCTTGGGCAATACATGAAAT 877  
Qy 239 ProThrGlnSerLysAlaThrGln---ProSerSerSerThrValProLysThrGluAsp 257  
Db 878 TCTGGCTTCAACAAAGTCTCTCCCTTGGAGGGGCACAAAGCATCAGTAAGAATACAG 937  
Qy 258 GlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsn 277  
Db 938 CAACGGCCCCAGCAGATCCGTATCAGATCTTGGCCCCGACGAGTGCCTAGCCAC 997  
Qy 278 ProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeuLeuSerSerSer 297  
Db 998 CTTGGAAGCGGCGAGATCCAGCTGTGGCAATCTCTCTCTGGAGTCTCTCCGACAGCGCC 1057  
Qy 298 AsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspProAsp 317  
Db 1058 AACGCCAGCTGTATCACCTGGAGGGGACCAACGGGGAGTTCAAAATGACGAGCCCGCAT 1117  
Qy 318 GluValAlaArgArgTyrGlyLysSerLysProAsnMetAsnTyrAspLysLeu 337  
Db 1118 GAGGTGGCCAGCGCTGGGGCGAGCGGAAAGCAAGCCCAACATGAATTACACAAGCTG 1177  
Qy 338 SerArgAlaLeuArgTyrTyrTrpAspLysAsnIleMetThrLysValHisGlyLysArg 357  
Db 1178 AGCCGGGCCCTCCGTATTACTATGATAAAACATTTATGACCAAGTGCACGGGCAAGA 1237  
Qy 358 TyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProProGlu 377  
Db 1238 TATGCTTACAATTTGACTTCCAGGCATTGCCAGGCTCTGCAGCCACATCCGACCGAG 1297  
Qy 378 SerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisPro 397  
Db 1298 TCGTCCATGTACAAGTACCTCTTGCATCTCTACATGCTCTTCTACATGCCACCAG 1357  
Qy 398 GlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSerSerSerPhe 417  
Db 1358 CAGAGGTGAACTTGTCTCCCTCCCATCTCTCCATCTCTCTCTCTCTCTCTCTCT 1417  
Qy 418 PheAlaAlaProAsnProTyrTrpAsnSerProThrGlyGlyIleTyrProAsn----- 435  
Db 1418 TTTGGAGCGCATCAACAATACATACTGGACCTCCCGCCAGCGGGGAATCTACCCCAACCCCAAC 1477  
Qy 436 ---ThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451  
Db 1478 GTCCCCCGCCATCTTACACCCAGCTGCTTACACATTTAGGAGCTACTAC 1528  
RESULT 14  
ABK84139  
ID ABK84139 standard; cDNA; 2957 BP.  
XX AC ABK84139;  
XX XX  
DT 14-AUG-2002 (first entry)  
XX XX  
DE Human cDNA differentially expressed in granulocytic cells #710.  
XX Human; ss; granulocytic cell; DNA chip; bacterial infection;  
KW viral infection; parasitic infection; protozoal infection;  
KW fungal infection; sterile inflammatory disease; psoriasis;  
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;  
KW adult respiratory distress syndrome; inflammatory bowel disease;  
KW Crohn's disease; ulcerative colitis; periodontal disease;  
KW granulocyte activation; chronic inflammation; allergy.  
XX Homo sapiens.  
XX WO200228999-A2.  
XX XX  
PD 11-APR-2002.







```
Db      1118 GAGGTGGCCAGCGCTGGGGGAGCGGAAAGCAAGCCCAACATGAAATTCAGCAAGCTG 1177
Qy      338 SerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArg 357
Db      1178 AGCCGGGCCCTCCGTTATTACTATGATAAAAAACATTATGACCAAAAGTGACGCGCAAAAGA 1237
Qy      358 TyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGlu 377
Db      1238 TATGCTTACAAATTGACTTCCAGCGCATTCGCCAGGCTGTGCAGCCACATCCGACCCGAG 1297
Qy      378 SerSerMetTyrLysTyrProSerAspLeuProTyrMetSerTyrHisAlaHisPro 397
Db      1298 TCGTCCATGTACAGTACCTTCTGACATCTCTTACATGCTTCTTACCATGCCACCCAG 1357
Qy      398 GlnLysMetAsnPheValAlaProHisProAlaLeuProValThrSerSerSerPhe 417
Db      1358 CAGAAGGTGAACCTTTGTCCCTCCCATCCATCCTCCATGCTGTCACCTTCTCCAGCTTC 1417
Qy      418 PheAlaAlaProAsnProTyrTyrAsnSerProThrGlyGlyIleTyrProAsn----- 435
Db      1418 TTTGAGCCGCATCACAATATCTGACCTCCCCCAGGGGGGAACTACCCCAACCCCAAC 1477
Qy      436 ---ThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
Db      1478 GTCCCCCGCCATCCTTAACACCCACGTCGCTTACACTTAGGCAGCTACTAC 1528
```

Search completed: May 27, 2004, 21:22:13  
Job time : 642 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 27, 2004, 20:59:17 ; Search time 128 Seconds  
(without alignments)  
1955.335 Million cell updates/sec

Title: US-09-902-772-2

Perfect score: 2444

Sequence: 1 MASTIKKALSVSDQSLFE.....LYPNTRLPAAHMPSHLGTYV 451

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US0902772/runat27052004\_162842\_2168/app\_query.fasta\_1.647  
-DB=Issued Patents NA -QMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US0902772@cgn 1 1 69 @runat27052004\_162842\_2168 -NCPU=6 -ICPU=3  
-NO WMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTCUTS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query %	Score	Match	Length	DB ID	Description
1	2444	100.0	1447	3	US-08-878-177-1	Sequence 1, Appli
2	2414.5	98.8	1528	3	US-08-878-177-3	Sequence 3, Appli
3	1624.5	66.5	2938	2	US-08-343-443B-3	Sequence 3, Appli
4	511	20.9	1752	3	US-09-350-779-1	Sequence 1, Appli
5	511	20.9	1752	4	US-09-435-335-1	Sequence 1, Appli
6	449	18.4	1604	1	US-08-306-691B-43	Sequence 43, Appli
7	449	18.4	1604	5	PCT-US93-06251-9	Sequence 9, Appli
8	445.5	18.2	2268	3	US-09-344-579-1	Sequence 1, Appli
9	407	16.7	2532	4	US-08-620-312D-869	Sequence 869, App
10	342.5	14.0	2667	2	US-08-459-412A-1	Sequence 1, Appli
11	342.5	14.0	2667	3	US-09-021-715-1	Sequence 1, Appli
12	322	13.2	2064	3	US-08-875-944B-1	Sequence 1, Appli

13	322	13.2	2064	3	US-09-116-049-3	Sequence 3, Appli
14	322	13.2	2064	4	US-09-602-868A-1	Sequence 1, Appli
15	322	13.2	2064	4	US-09-884-363-3	Sequence 3, Appli
16	306	12.5	2410	2	US-08-780-835B-1	Sequence 1, Appli
17	306	12.5	2410	3	US-08-303-268-1	Sequence 1, Appli
18	306	12.5	2410	3	US-09-116-049-1	Sequence 1, Appli
19	306	12.5	2410	4	US-09-884-363-1	Sequence 1, Appli
20	306	12.5	2410	4	US-09-328-525-13	Sequence 13, Appli
21	301.5	12.3	328	2	US-08-343-443B-5	Sequence 5, Appli
22	301	12.3	1894	4	US-09-570-593-1	Sequence 1, Appli
23	301	12.3	1905	3	US-09-055-113-2	Sequence 2, Appli
24	301	12.3	3317	4	US-09-570-593-12	Sequence 12, Appli
25	291.5	11.9	2266	2	US-08-213-767-1	Sequence 1, Appli
26	279.5	11.4	1933	4	US-09-920-759-3	Sequence 3, Appli
27	279.5	11.4	1976	4	US-09-920-759-10	Sequence 10, Appli
28	273.5	11.2	5510	3	US-09-009-913-3	Sequence 3, Appli
29	261.5	10.7	5427	3	US-09-009-913-2	Sequence 2, Appli
30	261.5	10.7	5667	3	US-09-009-913-4	Sequence 4, Appli
31	259	10.6	665	4	US-09-920-759-11	Sequence 11, Appli
32	234.5	9.6	2544	2	US-08-459-412A-6	Sequence 6, Appli
33	234.5	9.6	2544	3	US-09-021-715-6	Sequence 6, Appli
34	233	9.5	731	4	US-09-976-594-191	Sequence 191, App
35	231.5	9.5	848	3	US-09-009-913-338	Sequence 338, App
36	231.5	9.5	2280	3	US-09-009-913-8	Sequence 8, Appli
37	231.5	9.5	2428	3	US-09-009-913-6	Sequence 6, Appli
38	231.5	9.5	2498	3	US-09-009-913-10	Sequence 10, Appli
39	230	9.4	1907	4	US-09-300-958A-27	Sequence 27, Appli
40	230	9.4	1907	4	US-09-570-593-4	Sequence 4, Appli
C 41	229	9.4	852	3	US-09-020-566-44	Sequence 44, Appli
C 42	229	9.4	852	3	US-09-030-607-44	Sequence 44, Appli
C 43	229	9.4	852	3	US-09-439-313-44	Sequence 44, Appli
C 44	229	9.4	852	4	US-09-352-616A-44	Sequence 44, Appli
C 45	229	9.4	852	4	US-09-232-149A-44	Sequence 44, Appli

#### ALIGNMENTS

RESULT 1  
US-08-878-177-1  
; Sequence 1, Application US/08878177

; Patent No. 6294354

; GENERAL INFORMATION:

; APPLICANT: Chugai Seiyaku Kabushiki Kaisha, 5001, Iwamoto et

; TITLE OF INVENTION: Cell Calcification Suppressing Proteins and Genes of

; FILE REFERENCE: chugai seiyaku kabushiki kaisha 5001

; CURRENT APPLICATION NUMBER: US/08/878,177

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1

; LENGTH: 1447

; TYPE: DNA

; ORGANISM: C-11 gene, c-erg gene w/ deletion, chicken DNA

US-08-878-177-1

#### Alignment Scores:

Pred. No.: 2.54e-249 Length: 1447  
Score: 2444.00 Matches: 451  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 3 Gaps: 0

US-09-902-772-2 (1-451) x US-08-878-177-1 (1-1447)

QY 1 MetAlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20

63 ATGGCAGCACTATTATAGGAAGCATTTATCAGTGGTGGTGAAGACCCAGCTCTTTTGG 122

DB 21 CysAlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGlu 40

123 TGTGCGCTACGGATCGCCCACTTGCAAGACAGAAATGACAGCCCTCTCTCCATGAA 182

QY 41 TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnPro 60  
 DB 183 TATGGGCAAAACATCAAGAATGAGCCCGCGCTTCCCGACGAGACTGGTTATCAAGAGCC 242  
 QY 61 ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80  
 DB 243 CCGGCCAGAGTTACCATTAAGATGGAGTGTACCCCAACACAGGTTAATGGGTCAAGGAAT 302  
 QY 81 SerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsnVal 100  
 DB 303 TCACCTGATGACGTGCGCGTGGCAAGAGGAGGAAATGGTTAGCAGTTCAGACAAATGTT 362  
 QY 101 GlyMetAsnTyrGlySerTyrMetGluGluLysHisIleProProAsnMetThrThr 120  
 DB 363 GGGATGAACTATGGAGCTACATGGAGAGAGCAATATTCGGCTCCAAATATGACACC 422  
 QY 121 AsnGluArgValIleValProAlaAspProThrLeuTyrSerThrAspHisValArg 140  
 DB 423 AATCAACGAAGAGTTATTTGGCCAGAGATCTTACCGTTATGGAGCAGACACCATGTACGG 482  
 QY 141 GlnTyrLeuGluTyrAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPhe 160  
 DB 483 CAGTGGCTGGAGTGGGAGTGAAGAGTATGGTCTCCAGACGTGACATCTTGTGTTC 542  
 QY 161 GlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThr 180  
 DB 543 CAGAACATTCATGGGAAGAGTGTGTGTAATGACCAAGATGACATTCACAGAGACTCACG 602  
 QY 181 ProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArgGlyAla 200  
 DB 603 CCGAGCTATTAACGAGATATCTCCCTGTCACACCTACCTACCTCAGAGAGAGAGGCC 662  
 QY 201 ThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrArg 220  
 DB 663 ACTTTTATTTTCCAAATACATCAGTTTACCCAGAGCAACGCAAGAATAACAACAAGG 722  
 QY 221 ProAspLeuProTyrGluGlnAlaArgArgSerAlaTyrThrSerHisSerHisProThr 240  
 DB 723 CCAGATTACCTTATGAGCAAGAGAGAGATCAGCGTGGAGCTGACAGCCATCCCAT 782  
 QY 241 GlnSerLysAlaThrGlnProSerSerThrValProLysThrGluAspGlnArgPro 260  
 DB 783 CAGTCAAAAGCTACCCNACCATCATCTTCNACAGTGCCTCCCAACAGACACAGCGTCT 842  
 QY 261 GlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGlySer 280  
 DB 843 CAGTTAGATCCTTATCAGATTCTTGACCCAGCAGCGCTCTTGCAAAATCCAGGGAGT 902  
 QY 281 GlyGlnIleGlnLeuTyrGlnPheLeuLeuGluLeuSerAspSerSerAsnSerAsn 300  
 DB 903 GGGCAGATACAGTATGGCAGTTCCTACTGGAGCTTCTGTCGACAGCTCCAACTCCAC 962  
 QY 301 CysIleThrTyrGluGlyThrAsnGlyGluPheLysMetThrAspProAspGluValAla 320  
 DB 963 TGCATACCTGGGAGGCAAAATGGGAGTTCAGATGACAGACCCCTGTAAGTGGCT 1022  
 QY 321 ArgArgTyrGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArgAla 340  
 DB 1023 CGCGTGTGGGAGAGAGAAAGCAACCTTAACATGATCATGACAACTCAGCCGTGCA 1082  
 QY 341 LeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaTyr 360  
 DB 1083 CTTGCTACTTACTATGACAAAAATATTATGACTAAAGTTTCATGGTAAACGCTATGCTAC 1142  
 QY 361 LysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProProGluSerSerMet 380  
 DB 1143 AAATTTGATTTCCACGGAATCGCTCAGCGCCCTCAGCCCTCAGCAATCATCATG 1202  
 QY 381 TyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisProGlnLysMet 400  
 DB 1203 TACAAATACCATCAGACCTCCCTCATATGATGTTCTCTCATGACACCCCCAGAGATG 1262

QY 401 AsnPheValAlaProHisProHisProAlaLeuProValThrSerSerSerPheAlaAla 420  
 DB 1263 AACTTTGTAGTCCCATCCCATCCCTGCTTTCGCGTAACCTCACTCCAGCTTTTGTGTC 1322  
 QY 421 ProAsnProTyrTyrAsnSerProThrGlyGlyIleTyrProAsnThrArgLeuProAla 440  
 DB 1323 CCTAATCCATCTGGAATTCACCACTGGAGGATCTACCCCAATACCAAGCTGCCAGCT 1382  
 QY 441 AlaHisMetProSerHisLeuGlyThrTyrTyr 451  
 DB 1383 GCTCATATGCTTCCCATCTTGGCACCTACTAC 1415  
 RESULT 2  
 US-08-878-177-3  
 ; Sequence 3, Application US/08878177  
 ; Patent No. 6294354  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Chugai Seiyaku Kabushiki Kaisha, 5001, Iwanoto et  
 ; TITLE OF INVENTION: Cell Calcification Suppressing Proteins and Genes of  
 ; FILE OF INVENTION: the Proteins  
 ; FILE REFERENCE: chugai seiyaku kabushiki kaisha 5001  
 ; CURRENT APPLICATION NUMBER: US/08/878.177  
 ; CURRENT FILING DATE: 1997-06-18  
 ; NUMBER OF SEQ ID NOS: 7  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 3  
 ; LENGTH: 1528  
 ; TYPE: DNA  
 ; ORGANISM: c-erg gene, chicken DNA  
 US-08-878-177-3  
 Alignment Scores:  
 Pred. No.: 3,74e-246 Length: 1528  
 Score: 2414.50 Matches: 450  
 Percent Similarity: 94.14% Conservative: 0  
 Best Local Similarity: 94.14% Mismatches: 1  
 Query Match: 98.79% Indels: 27  
 DB: 3 Gaps: 1  
 US-09-902-772-2 (1-451) x US-08-878-177-3 (1-1528)  
 QY 1 MetAlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20  
 DB 63 ATGCGAAGCACTATTAAAGGAAGCATTTATCAGTGTGTGAGTGAAGACCAAGTCTCTTTGAG 122  
 QY 21 CysAlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGlu 40  
 DB 123 TGTGCTACGGATCGCCCATCTTGCAAGAGACAGAAATGACAGCTCTCTTCCAGTGA 182  
 QY 41 TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnPro 60  
 DB 183 TATGGCAAAACATCAAGATGAGCCCGCGCTTCCCGACGAGGACTGGTTATCACAGCCC 242  
 QY 61 ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80  
 DB 243 CCGGCCAGAGTTACCATTAAGATGGAGTGTAAACCAACCAAGGTTAATGGTCAAGGAAT 302  
 QY 81 SerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsnVal 100  
 DB 303 TCACCTGATGATCGAGCTGCGCAAGAGGAGGAAATGGTTAGCAGTTCAGACAAATGTT 362  
 QY 101 GlyMetAsnTyrGlySerTyrMetGluGluLysHisIleProProAsnMetThrThr 120  
 DB 363 GGGATGAACTATGGAAAGCTACATGGAAGAGAGATATTCGCGCTCCAAATATGACACC 422  
 QY 121 AsnGluArgValIleValProAlaAspProThrLeuTyrSerThrAspHisValArg 140  
 DB 423 AATGAACGAAGAGTTATTTGTCGAGCAGATCTTACGTTATGGAGCAGACACCATGTACGG 482  
 QY 141 GlnThrLeuGluTyrAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPhe 160  
 DB 483 CAGTGGCTGGAGTGGGAGTGAAGAGTATGGTCTCTCCAGACGTGACATCTTGTGTTC 542

```

QY 161 GlnAsnLeuaspGlyLysGluLeuCysLysMetThrLysAspPheGlnArgLeuThr 180
Db 543 CAGAACATTGATGGAAAGATTGTGTAAATGACCAAGATGACTTCCAGAGACTCAGC 602
QY 181 ProSerTyArgAlaAspIleLeuLeuSerHisLeuHisTyArgGluArg----- 198
Db 603 CCGAGCTATAACGAGATATCTCTGTGCACACCTACCTACCTACCTACCTACCTCTT 662
QY 198 ----- 198
Db 663 CCACATTGACTTCAGATGATGTTGATAAGCCCTTACAAACTCTCCACGGTTAATGCAT 722
QY 199 -----GlyAlaThrPheIlePheProAsnThrSerValTyProGluAla 213
Db 723 GCTAGAAACACAGAGAGGCCATTTTATTTTCCAAATACATCAGTTTACCAGAGCA 782
QY 214 ThrGlnArgIleThrArgProAspLeuProTyArgGlnAlaArgArgSerAlaTyr 233
Db 783 ACGCAAGAAATAACAAAGCCAGATTACCTTATGAGCAAGCGAGAGATCAGCGTGG 842
QY 234 ThrSerHisSerHisProThrGlnSerLysAlaThrGlnProSerSerThrValPro 253
Db 843 ACGAGTCACGCCATCCCATCTCATGTCAGTCAAGCTACCCCAACATCTTCAACAGTCCC 902
QY 254 LysThrGluAspGlnArgProGlnLeuAspProTyArgGlnIleLeuGlyProThrSerSer 273
Db 903 AAAACAGAGACCAAGCGTCTCAGTTAGATCTTATCAGATCTTGTGACCGACGAGCAGC 962
QY 274 ArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuThrGlnPheLeuGluLeu 293
Db 963 CGTCTTGCAATCCAGGAGTGGGAGATACAGTATGGCAGTTCTTACTGGAGCTTCGT 1022
QY 294 SerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMet 313
Db 1023 TCGACAGCTCCACTCCACTGCATCACCTGGGAGGCAAAATGGGAGTTCAAGATG 1082
QY 314 ThrAspProAspGluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMetAsn 333
Db 1083 ACAGACCTGTGATGAGTGGCTCGCGTGGGAGAGAGAGAAAGCAACCTAACATGAAC 1142
QY 334 TyrAspLysLeuSerArgAlaLeuArgTyTrpTyArgLysAspLysAsnIleMetThrLysVal 353
Db 1143 TATGACAAACTCAGCGTGCCTTCTGCTACTACTATGACAAAATATATATGACTAAGTT 1202
QY 354 HisGlyLysArgTyArgAlaTyLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnPro 373
Db 1203 CATGTAACGCTATGCTCAAAATTTGATTTCCAGCGAATCGCTCAGGCGCTCCAGCT 1262
QY 374 HisProGluSerSerMetTyLysTyTrpSerAspLeuProTyTrpMetSerSerTyr 393
Db 1263 CACCTCCAGAAATCATCATGTAATAATCCCATCAGACCTCCCTCCATGAGTTCTCTAC 1322
QY 394 HisAlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProValThr 413
Db 1323 CATGCACACCCCGCAGACAGTGAATTTGTAGCTCCCTCCATCCCTGCTTTGCGCGTAACC 1382
QY 414 SerSerSerPhePheAlaAlaProAsnProTyTrpAsnSerProThrGlyGlyIleTyr 433
Db 1383 TCATCCAGCTTTTGTGTCGCCCTTAATCCATCTAGTGAATCCCACTGGAGGATCTAC 1442
QY 434 ProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyTrp 451
Db 1443 CCCAATACAGGCTGCCAGCTGCATATGCTTCCCATCTTGGCACCTACTAC 1496

```

RESULT 3

```

US-08-343-443B-3
; Sequence 3, Application US/08343443B
; Patent No. 5968734
; GENERAL INFORMATION:
; APPLICANT: Aurias, Alain
; APPLICANT: Delattre, Olivier
; APPLICANT: Desmaze, Chantal
; APPLICANT: Melot, Thomas

```

```

; APPLICANT: Peter, Martine
; APPLICANT: Ploougaestel, Beatrice
; APPLICANT: Thomas, Gilles
; APPLICANT: Zucman, Jessica
; TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF
; TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL
; TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS
; TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID
; TITLE OF INVENTION: TRANSLOCATIONS
; NUMBER OF SEQUENCES: 129
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: AEDIT 1.0 DOS text editor
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/343,443B
; FILING DATE: 18-NOV-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR93/00494
; FILING DATE: 19-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/06123
; FILING DATE: 20-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 989,6121P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2938 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 143..1498
; US-08-343-443B-3

```

Alignment Scores:

Pred. No.:	Score:	Length:
6,89e-162	1624.50	2938
Percent Similarity:	79.43%	Matches: 306
Best Local Similarity:	66.96%	Conservative: 57
Query Match:	66.47%	Mismatches: 83
		Indels: 11
		Gaps: 7

US-09-902-772-2 (1-451) x US-08-343-443B-3 (1-2938)

```

QY 1 MetAlaSerThrLysLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20
Db 143 ATGGACGGGACTATTAAAGGAGGCTCTGTCGGTGGTGGAGCGAGCAGCCAGTCCCTCTTTGAC 202
QY 21 CysAlaTyGly---SerProHisLeuAlaLysThrGluMetThrAlaSerSerSer 39
Db 203 TCAGCGTACGAGCGCGGACGCCCATCTCCCCAGCGCGGACATGACTCTCTCGGGGAGTCT 262
QY 40 GluTyArgLysGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGln 59
Db 263 GACTACGGGCGAGCCGCCACAGATCAACCCCTCCACACAGAGAGTGGATCAATCAG 322
QY 60 ProProAlaArgValThrLysMetGluCysAsnProAsnGlnValAsnGlySerArg 79

```

Db 323 ---CGAGTGGGGTCAAGCTCAAGCGGAGTAT-----GACCACATGATGATGATCCAGG 373  
Qy 80 AsnSerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsn 99  
Db 374 GAGTCTCCGGTGGAGTCCAGCTGAGCAATGCAAGCTGGTGGCGGAGCGAGTCC 433  
Qy 100 ValGlyMetAsnTyrGlySerTyrMetGluGluLysHis---1leProProAsnMet 118  
Db 434 AACCCCATGAACTACACAGCTATATGACGAGAGAAATGGCCCCCTCTCTCCCAATG 493  
Qy 119 ThrThrAsnGluAArgValIleValProAlaAspProThrLeuTyrSerThrAspHis 138  
Db 494 ACCACCAACGAGAGAGAGTATCGTCCCGCAGACCCACACTGTGTGCACACAGAGCAT 553  
Qy 139 ValArgGlnTyrPLeuGluTyrAlaValLysGluTyrGlyLeuProAspValAspIleLeu 158  
Db 554 GTGAGCAATGGCTGGAGTGGCCATAAAGAGATATAGCTTGTATGAGATCGACATCC 613  
Qy 159 LeuPheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArg 178  
Db 614 TTTTTCAGACATCGATGGCAAGAACTGTGTAAATGAACAAGAGAGGACTTCTCCCGC 673  
Qy 179 LeuThrProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArg 198  
Db 674 GCCACACCTCTACAAACGCAAGTGTGTGTACACCTCAGTTACCTCAGGGAA--- 730  
Qy 199 GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThr 218  
Db 731 ---AGTTCAGCTGCTGCTATATAACACCTCCACACAGCAATCTCTCAGGATGAGT 787  
Qy 219 ThrArgProAspLeuProTyrGluGlnAlaArgSerAlaTyrThrSerHisSerHis 238  
Db 788 GTCAAAAGAACCTCTTTATGATCATGTCAGAGAGAGCATGGGCAATAACATGAT 847  
Qy 239 ProThrGlnSerLysAlaThrGln---ProSerSerSerThrValProLysThrGluAsp 257  
Db 848 TCTGGCTCAACAAAAGTCTCTCCCTTGGAGGGGCACAAACGATCAGTAAGATACAGAG 907  
Qy 258 GlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsn 277  
Db 908 CAACGGCCCCAGCAGATCCGATACAGATCTGGGCCCGCACAGCATCGCTTAGCCAAC 967  
Qy 278 ProGlySerGlyGlnIleGlnLeuTyrGlnPheLeuLeuGluLeuLeuSerAspSerSer 297  
Db 968 CTTGGAGCGGCGAGATCCAGCTGTGGCAATCTCTCTGAGGCTGCTCTCGACAGGCC 1027  
Qy 298 AsnSerAsnCysIleThrTyrGluGlyThrAsnGlyGluPheLysMetThrAspProAsp 317  
Db 1028 AACGCCAGCTATACCTGGAGGGGACCAACGGGAGTTCAAAATGACGGACCCCGAT 1087  
Qy 318 GluValAlaArgAspTyrGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeu 337  
Db 1088 GAGTGGCCAGCGCTGGGGCGAGCGGAAAGCAAGCCCAACATGAAATACGACAGCTG 1147  
Qy 338 SerArgAlaLeuAArgTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArg 357  
Db 1148 AGCGGGGCCCTCCGTATTATATGATAAAACATTTATGACCAAAAGTGCACGGCAAAAGA 1207  
Qy 358 TyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGlu 377  
Db 1208 TATGCTTACAAATTTGATCTCCAGCGCATTCGACGCTCTGACGACCATCCGACCGAG 1267  
Qy 378 SerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisPro 397  
Db 1268 TCGTCCATGTACAAAGTACCTCTGACATCTCTACATGCTCTCTACCATGCCACAG 1327  
Qy 398 GlnLysMetAsnPheValAlaProHisProAlaLeuProValThrSerSerSerPhe 417  
Db 1328 CAGAAGGTGAATTTGTCTCCCTCCCATCTCCATCTCCATGCTGTCATCTCTCCAGCTTC 1387  
Qy 418 PheAlaAlaProAsnProTyrTyrAsnSerProThrGlyGlyIleTyrProAsn----- 435

Db 1388 TTTGAGCGGCATCAATACTGGACCTCCCCACGGGGGAATCTACCCCAACCCCAAC 1447  
Qy 436 ---ThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451  
Db 1448 GTCCCCCGCCATCTTAACACCCACGTCCTTACACTTAGCAGCTACTAC 1498  
RESULT 4  
US-09-360-779-1  
; Sequence 1, Application US/09360779  
; Patent No. 6266216  
; GENERAL INFORMATION:  
; APPLICANT: Deneris, Evan S. V.  
; APPLICANT: Fyodorov, Dmitry V.  
; APPLICANT: Hendricks, Timothy J.  
; TITLE OF INVENTION: Reagents and Methods for the Screening of Compounds  
; TITLE OF INVENTION: Useful in the Treatment of Neurological Diseases  
; FILE REFERENCE: CASE-03828  
; CURRENT APPLICATION NUMBER: US/09/360,779  
; CURRENT FILING DATE: 1999-07-26  
; EARLIER APPLICATION NUMBER: 60/094,264  
; EARLIER FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 1752  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (112)..(1131)  
US-09-360-779-1  
Alignment Scores:  
Pred. No.: 3,849-44 Length: 1752  
Score: 511.00 Matches: 123  
Percent Similarity: 50.69% Conservative: 23  
Best Local Similarity: 42.71% Mismatches: 63  
Query Match: 20.91% Indels: 79  
Gaps: 10  
US-09-902-772-2 (1-451) x US-09-360-779-1 (1-1752)  
Qy 209 ValTyrProGluAlaThrGlnArgIleThrArgProAspLeuProTyrGluGlnAla 228  
Db 369 GTGCCCGCCCGGCGGTCCTCCCGCATCTCCACGCCCA----- 407  
Qy 229 ArgArgSerAlaTyrThrSerHisSerHisProThrGlnSerLysAlaThrGlnProSer 248  
Db 408 -----GTCCCGCGCAGCATGAGACGAGCGGCACCTCCAGCCCTGCT 452  
Qy 249 SerSerThrValProLysThrGluAspGlnArg-----Pro 260  
Db 453 GATCAACATGTACTTACCATCCGATCCGAGATGCTCTTTTAAGGAAGGAGAGGCC 512  
Qy 261 GlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGlySe 280  
Db 513 GAGCTGGGGGCGG-----CTGAGCCCTCGGTACAGAAA-----GGCAG 551  
Qy 280 rGlyGlnIleGlnLeuTyrGlnPheLeuLeuLeuLeuSerAspSerSerAsnSerAs 300  
Db 552 CGGGCGAGATCCAGTTGTGGCAGTTTCTACTGAGCTGTGGCAGACCGCGGACGCCGG 611  
Qy 300 nCysIleThrTyrGluGlyThrAsnGlyGluPheLysMetThrAspProAspGluVal 320  
Db 612 CTGCATCGCTGGAGGGCGCGCAGCGGAGTTCAAGCTCACCAGCCCGCAGCGGTGGC 671  
Qy 320 aArgArgTyrGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArgAl 340  
Db 672 GCGAGCTGGGCGGAGCGCAGAGCAAGCCCAATATGACTACGACAGCTAAGTCGAGC 731  
Qy 340 aLeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaTy 360  
Db 732 ACTGGGCTACTACTACGACAAACATCATGAGCAAGGTGTCAGCGCAAGCGCTACGCCCTA 791



```

; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/305,691B
; FILING DATE: September 15, 1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Monaco, Daniel A.
; REGISTRATION NUMBER: 30,480
; REFERENCE/DOCKET NUMBER: 8321-8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-8383
; TELEFAX: (215) 568-5549
; TELEX: No. 5734039e
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1604 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-306-691B-43

Alignment Scores:
Pred. No.: 1,25e-37 Length: 1604
Score: 449.00 Matches: 114
Percent Similarity: 43.88% Conservative: 51
Best Local Similarity: 30.32% Mismatches: 89
Query Match: 18.37% Indels: 122
DB: 1 Gaps: 9

US-09-902-772-2 (1-451) x US-08-306-691B-43 (1-1604)

Qy 119 ThrThrAsnGluArgValIleValProAlaAspProThrLeuTrpSerThrAspHis 138
Db 447 ACTAAGAACAGCAACGACTGGGGATCCCAAGAGACCCCGGCGAGTGGACAGAAACCCAT 506
Qy 139 ValArgGlnTrpLeuGluTrpAlaValLysGluTrpGlyLeuProAspValIleLeu 158
Db 507 GTTCGGGACTGGGGTGTGGGTGTGAATTCAGCTGAAAGTGTAGACTTCCAG 566
Qy 159 LeuPheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArg 178
Db 567 AAGTTC---TGATGATGAGCAGAGCCCTCTGGCCCTGGTAAAGACTGCTTTCTCGAG 623
Qy 179 LeuThrProSerTrpAsnAlaAspIleLeuLeuSerHisLeuHisTrpLeuArgGlu--- 197
Db 624 CTGCCCCCAGACTTTGTTGGGACATCTTATGGGAACATCTAGAGATCCTGCGAAGAG 683
Qy 198 -----ArgGlyAlaThrPheIlePheProAsnThr----- 207
Db 684 GATGTGAACCATATCAAGTTAATGGAGTCAACCCAGCTATCCAGATCCCGCTATACC 743
Qy 207 ----- 207
Db 744 TCGGATTACTTCAATAGCTATGTTAGTATGAGCATGCCAGTGTGTTCACCATCCGGAGTTC 803
Qy 208 -----SerValTrpProGluAlaThrGlnArg 216
Db 804 TCAGAGCCAGCTTCATCAGAGTCTCATCAGCGCTCCATCCCATCAGCTCGGAGAG 863
Qy 217 Ile----- 217
Db 864 CTCCTCTCCCTCAAGTATGAGATGACTACCCCTCGGTCTATTTCTCCGAGACCCCTCTCCAG 923
Qy 217 ----- 217
Db 924 ACAGACACCTTGCAGATGACTACTTTGCTATCAAAACAAGAGTCTGCACCCAGACAAC 983
Qy 218 -----ThrThrArgProAspLeuProTrpGluGlnAlaArgArgSer--- 231
Db 984 ATGTGCATGGGGAGGACCGAGTCGTGGTAAACTCGGGGGCCAGGACTCTTTTGAAGACATA 1043

```

```

Qy 232 ---AlaTrpThrSerHisSerHisProThrGlnSerLysAlaThrGlnProSerSerSer 250
Db 1044 GAGAGCTACGATAGTGTGATCGCTCACCAGCTCTGGAGCAGCCAGTCACTTTTCAAC 1103
Qy 251 ThrValProLys-----ThrGluAsp----- 257
Db 1104 AGCCTGACGGTGTTCCTCTCTATGACAGCTTGACACTCAGAGGACTATCGGGTGCCTG 1163
Qy 258 -----GlnArgProGlnLeuAsp 263
Db 1164 CCCAACCAAGCCCAAGGCGACCTTCAAGGACTATGTGGGGACCTGTGCTACCTCAAT 1223
Qy 264 ProTrpGlnIleLeuGlyProThrSerSerArgLeuAlaSerProGlySerGlyGlnIle 283
Db 1224 AAGGACAAAGCCTGTCTTCCTCTGCTGCTGCTAGCTGGCTACACAGCAGTGGACCAATC 1283
Qy 284 GlnLeuTrpGlnPheLeuLeuGluLeuLeuSerAspSerSerAsnSerAsnCysIleThr 303
Db 1284 CAGCTATGGCAGTTCTTCTTGGAAATTAATCTACTGATAAATCCTCTCAGTCTTTATCAGC 1343
Qy 304 TrpGluGlyThrAsnGlyGluPheLysMetThrAspProAspGluValAlaArgArgTrp 323
Db 1344 TGGACAGGAGATGGCTGGGAATTCAAACTTCTGACCCAGATGAGTGGCCAGAGATGG 1403
Qy 324 GlyGluArgLysSerLysProAsnMetAsnTrpAspLysLysLeuSerArgAlaLeuArgTrp 343
Db 1404 GGAAGAGGAAAAACAAACCTTAAGATGAATATGAGAACTTGAGCCGTGGCTACGCTAC 1463
Qy 344 TrpTrpAspLysAsnIleMetThrLysValHisGlyLysArgTrpAlaTrpLysPhe--- 362
Db 1464 TATTACGACAAAAACATCATCCACAGACAGCGGGGAAACGCTACGTGTACCGCTTTGTG 1523
Qy 363 ---AspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGlu 377
Db 1524 TGTGACCTG-----CAGAGCTGTGGGTACACCCCTGAG 1559

RESULT 7
PCT-US93-06251-9
; Sequence 2, Application PC/TUS9306251
; GENERAL INFORMATION:
; APPLICANT: Wickstrom, Eric and Rife, Jason P.
; TITLE OF INVENTION: Trivalent Synthesis of Oligonucleotides Containing
; TITLE OF INVENTION: Stereospecific Alkylphosphonates and Arylphosphonates
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: NY
; COUNTRY: USA
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/06251
; FILING DATE: 19930630
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 8586
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343
; TELEFAX: 516-742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1604 base pairs
; TYPE: nucleic acid

```

```
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
PCT-US93-06251-9

Alignment Scores:
Pred. No.: 1,25e-37 Length: 1604
Score: 449.00 Matches: 114
Percent Similarity: 43.88% Conservative: 51
Best Local Similarity: 30.32% Mismatches: 89
Query Match: 18.37% Indels: 122
DB: 5 Gaps: 9

US-09-902-772-2 (1-451) x PCT-US93-06251-9 (1-1604)
QY 119 ThrThrAsnGlnArgValIleValProAlaAspProThrLeuTrpSerThrAspHis 138
DB 447 ACTAAGACAGACAGACGACTGGGATCCCAAGAGACCCCGGAGTGGACAGAACCCAT 506
QY 139 ValArgGlnTrpLeuGlnTrpAlaValLysGluTrpGlyLeuProAspValAspLeu 158
DB 507 GTTCGGGACTGGGTGATGTGGCTGTGAATGAATTCAGCCTGAAGAGTGTAGACTTCCAG 566
QY 159 LeuPheGlnAsnLeuAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArg 178
DB 567 AAGTTC---TGATGAATGGAGCAGCCCTCTGGCCCTGGGTAAAGACTGCTTCTCGAG 623
QY 179 LeuThrProSerTyAsnAlaAspIleLeuLeuSerHisLeuHisTyLysArgGlu--- 197
DB 624 CTGGCCCCAGACTTGTGTGGGACATCTTATGGAAACATCTAGAGATCTCTGCAGAAAGAG 683
QY 198 -----ArgGlyAlaThrPheLeuPheProAsnThr----- 207
DB 684 GATGTGAACCATATCAAGTTAATGGAGTCAACCCAGCCTATCCAGATCCCGCTATACC 743
QY 207 ----- 207
DB 744 TCGGATTACTTTCATAGCTATGTTATGAGCATGCCCGCTGTCTCCACCATCGGAGTTC 803
QY 208 -----SerValTyProGluAlaThrGlnArg 216
DB 804 TCAGAGCCAGCTTCATCAGAGATCCTATCAGAGCTCCATCCCATCAGCTCGGAAGAG 863
QY 217 Ile----- 217
DB 864 CTCCTCTCCCTCAAGTATGAGATGACTACCCCTCGGTCAATCTCCGAGACCTCTCCAG 923
QY 217 ----- 217
DB 924 ACAGACACCTTCAGAGATGACTACTTTGTCTATCAACAGAGTCGTCACCCAGACAAC 983
QY 218 -----ThrThrArgProAspLeuProTyArgGlnAlaArgAspSer--- 231
DB 984 ATGTGCATGGGAGGACAGTCGTGGTAACTCGGGGCCGAGGACTCTTTTGAAGCATA 1043
QY 232 ---AlaTrpThrSerHisSerHisProThrGlnSerLysAlaThrGlnProSerSer 250
DB 1044 GAGAGCTACGATAGTGTGTGCGCTCACCAGTCCCTGGAGCAGCCAGTCACTCTTCAAC 1103
QY 251 ThrValProLys-----ThrGluAsp----- 257
DB 1104 AGCCTGACGGGTGTCCTCCTATGACAGCTTCGACTCAGAGGACTATCGGCTGCCTG 1163
QY 258 -----GlnArgProGlnLeuAsp 263
DB 1164 CCCAACACAGCCCAAGGACCTTCAGGACTATGTGGGACCGCTGTGACCTCAAT 1223
QY 264 ProTyGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGlySerGlyGlnIle 283
DB 1224 AAGGACAGCGCTGCTATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1283
QY 284 GlnLeuTrpGlnPheLeuLeuLeuLeuSerSerSerSerSerSerSerSerSerSer 303
DB 1284 CAGTATGGCAGTCTTCTTCTGGAAATTACTCAGTAAATCTCTCAGTCTTTTATCAGC 1343
QY 304 TrpGluGlyThrAsnGlyGluPheLysMetThrAspProAspGluValAlaArgArgTrp 323
DB 1344 TGGACAGAGATGGCTGGGAATTCANAATTTCTGACCCAGATGAGTGGCCAGGAGATGG 1403
QY 324 GlyGluArgLysSerLysProAsnMetAsnTyAspLysLeuSerArgAlaLeuArgTy 343
DB 1404 GGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1463
QY 344 TyTrpAspLysAsnIleMetThrLysValHisGlyLysArgTyAlaTyLysPhe--- 362
DB 1464 TATTACGCAAAACATCATCTCCACAGACCGGGGAAACGCTACGTACCGCTTGTG 1523
QY 363 ---AspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGlu 377
DB 1524 TGTGACCTG-----CAGAGCCTGCTGGGTACACCCCTGAG 1559

RESULT 8
US-09-344-579-1
; Sequence 1, Application US/09344579
; Patent No. 6054316
; GENERAL INFORMATION:
; APPLICANT: Brenda F. Baker
; APPLICANT: Lex M. Cowser
; TITLE OF INVENTION: ANTISENSE MODULATION OF ETS-2 EXPRESSION
; FILE REFERENCE: RTS-0063
; CURRENT APPLICATION NUMBER: US/09/344,579
; CURRENT FILING DATE: 1999-06-25
; NUMBER OF SEQ ID NOS: 47
; SEQ ID NO 1
; LENGTH: 2268
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (291)..(1700)
US-09-344-579-1

Alignment Scores:
Pred. No.: 5.2e-37 Length: 2268
Score: 445.50 Matches: 127
Percent Similarity: 40.96% Conservative: 43
Best Local Similarity: 30.60% Mismatches: 105
Query Match: 18.23% Indels: 141
DB: 3 Gaps: 14

US-09-902-772-2 (1-451) x US-09-344-579-1 (1-2268)
QY 122 GluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGln 141
DB 570 CAGCGCGCTGGGCAATTCAGAAAGAACCCCTGGCTGTGGAGTGAGCAACAGGTATGCCAG 629
QY 142 TrpLeuGluTrpAlaValLysGluTyArgGlyLeuProAspValAspIleLeuPheGln 161
DB 630 TGGCTTCTTGGGCCCAACCAATGATTCAGTCTGGTGAACGTGAATCTGCAGAGTTC--- 686
QY 162 AsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThrPro 181
DB 687 GCATGAATGGCCAGATGCTGTGTAACTTGGCAAGGAACGCTTCTGGAGCTGCACCT 746
QY 182 SerTyAsnAlaAspIleLeuLeuSerHisLeuHisTyLysArgGluArgGlyAlaThr 201
DB 747 GACTTGTGGTGACATTCCTGGGAACATCTGGGAACATCTGGGAACATCTCAAGAAACCAA--- 803
QY 202 PheIlePheProAsnThrSerValTyProGluAlaThrGlnArgIleThrThrArgPro 221
DB 804 -----GAAAGACAGAGATCAATATGAAGAAATTCACACTCACCTCCCTTCT 854
QY 222 -----AspLeuProTyArgGlnAla 228
DB 855 CATTGGATTACAGCAATACATTAGGTTTGGCAGACAGCGGCCCTTATGGAATGCAG 914
```



```

QY 228 aArgArg-----SerAlaThrThrSer-----HisSerHisProTh 240
Db 915 ACACAGAAATACCCAAAGCGGCTCTGGACAGCATGTGTCCGGCTCCACACCCAGC 974
QY 240 rGlnSerIysAlaThrGln-----ProSerSe 249
Db 975 GTACTCAGCTCTGACGAGGAGTTTCAGATGTTCCCAAGTCTCGGCTCAGCTCGTCCAGC 1034
QY 249 rSerThr----- 251
Db 1035 GTCACTACTGCTCTGCTCAGTCAGCACTTCCAGCAGCACTTGAATTGCTCACCAC 1094
QY 252 -----ValProLysThr-----GluAs 257
Db 1095 AATTCTGGGACTCCCAAGACACAGCACTCCCTCGAGAACGGTGGGACAGCTTCGAGAGC 1154
QY 257 pGlnArgProGlnLeuAspProTyrGlnIleLeuGlyPro----- 270
Db 1155 TCAGACTCCCTCTCCAGCTCTGGAACAGCAGCTGCTGCTGGATGTGCAACGGTT 1214
QY 270 ----- 270
Db 1215 CTTCTCTCGAGAGCTTCCAGAGATGACTGCAGCCAGTCTCTCTGCTCAATAAGCCAACC 1274
QY 271 -----ThrSerSerArgLeuAlaAsnPro----- 278
Db 1275 ATGTCTTTCAGGATTATCATCAAGAGAGG-AGTGACCCAGCTGGAGCAAGCAACAGT 1333
QY 279 -----GlySerGlyGlnIleGlnLeuLeuTrpGlnPh 288
Db 1334 TATACCTGCAGCTGTGCTGGCGGCTTTCACAGGAAGTGGACCTATTTCAGCTGTGGCAGTT 1393
QY 288 eLeuLeuGluLeuLeuSerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAs 308
Db 1394 TCTCTCGAGCTGCTATCAGCAAAATCTGCAGCTATTCATCAGCTGGAGCTGGAGCGG 1453
QY 308 nGlyLeuPheLysMetThrAspProAspGluValAlaArgTrpGlyGluArgLysSe 328
Db 1454 ATGGGAGTTTAACTCTGCGCCGACCCCGATGAGTGGCGCGCGGGTGGGAAAGAGGAAAA 1513
QY 328 rIysProAsnMetAsnTyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAs 348
Db 1514 TAAAGCCCAAGATGAATCAGAGAGCTGAGCGGGGTTTACGTACTATTTCAGCAAGAA 1573
QY 348 nIleMetThrLysValHisGlyLysArgTyrAlaTyrLysPhe----- 362
Db 1574 CATCATCCACAGAGCTCGGGGAAGCGCTAGCTGTACCGTTCGTGTGCGACCTCCAGAA 1633
QY 363 -----AspPheHisGlyIleAlaGlnAlaLeuGlnProHisPr 375
Db 1634 CTTGCTGGGTTTCACGCCGAGGAAGTACGACCCATCTG---GGCGTCCAGCCGACAC 1690
QY 375 oProGluSerSerMetTyrLysTyrProSerAspLeuPro---TyrMetSerSerTyrHi 394
Db 1691 GGAGGACTGAGGTGCGCGGGACCACTCCCTGAGCCGCGCGGCTCGTGGACTGAGTGGGA 1750
QY 394 sAlaHisProGlnLysMetAsnPheValAlaProHisProPro 408
Db 1751 AGCCCATCTCGACAGCTG-----CTCCCG 1775

```

## RESULT 9

```

US-09-620-312D-869
; Sequence 869, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom

```

```

; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; FILE REFERENCE: Polypeptides
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pc_FL_genes Version 1.0
; SEQ ID NO 869
; LENGTH: 2532
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2532)
US-09-620-312D-869

```

## Alignment Scores:

```

Pred. No.: 7,56e-33 Length: 2532
Score: 407.00 Matches: 94
Percent Similarity: 66.67% Conservative: 16
Best Local Similarity: 56.97% Mismatches: 23
Query Match: 16.65% Indels: 34
DB: 4 Gaps: 5

```

US-09-902-772-2 (1-451) x US-09-620-312D-869 (1-2532)

```

QY 271 ThrSerSerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeu 290
Db 2073 ACCCTCGAAGAGAAAGAAATGTCATGCGGGACACATCCAGCTGGGCGAGTTTCGCTG 2132
QY 291 GluLeuLeuSerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGlu 310
Db 2133 GAGCTGCTGCTGACCGCGGAACCGCGC-TGCATCGCTGGAGGGCGGTCAAGCGCAG 2191
QY 311 PheLysMetThrAspProAspGluValAlaArgArgTrpGlyGluArgLysSerLysPro 330
Db 2192 TTCAGCTCCAGACCGGACGAGGTGGCGCGCGGTGGCGGAGCGCAAGAGCAAGCCC 2251
QY 331 AsnMetAsnTyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMet 350
Db 2252 AACATGAATACGACAAAGCTGAGCCGCGCTGCT-ATCTACTACGACAAAGAACATCATG 2310
QY 351 ThrLysValHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAla 370
Db 2311 AGCAAGTGTCATGTAAGCGCTACGCTACCGCTTCGACTTCAGGGCTCGCGCAG--- 2367
QY 371 LeuGlnProHisProGluSerSerMetTyrLysTyrProSerSerAspLeu----- 387
Db 2368 -----GACGCGCGCTCTACAAAGCTGCCCGCGCGCTCCAGCCGCTG 2409
QY 388 -----ProTyrMetSerSerTyrHisAlaHisProGlnLysMetAsnPheValAlaPro 405
Db 2410 CCCTTCCCGGCTCTCC-----AACTCAACTCATGCGC--- 2445
QY 406 HisProProAlaLeuProValThrSerSerSerPhePheAlaAlaProAsnProTyrTrp 425
Db 2446 -----GCTCGCTGGAGGTGCGCGCGCGCT----- 2472
QY 426 AsnSerProThrGly 430

```

```

Db 2473 ---TCTCTACTGGC 2484
RESULT 10
US-08-469-412A-1
; Sequence 1, Application US/08469412A
; Patent No. 5856125
; GENERAL INFORMATION:
; APPLICANT: Mavrothalassitis, George J.
; APPLICANT: Blair, Donald G.
; APPLICANT: Fisher, Robert J.
; APPLICANT: Beal Jr., Gregory J.
; APPLICANT: Athanasios, Meropi A.
; APPLICANT: Sgouras, Dionysios N.
; TITLE OF INVENTION: The ERF Genetic Locus and Its Products
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,412A
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Garrett-Wackowski, Eugenia
; REGISTRATION NUMBER: 37,330
; REFERENCE/DOCKET NUMBER: 015280-229000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2667 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 123..1769
; OTHER INFORMATION: /note="human ERF (ETS2 Repressor
; OTHER INFORMATION: Factor) cDNA"
US-08-469-412A-1
Alignment Scores:
Pred. No.: 5,72e-26 Length: 2667
Score: 342.50 Matches: 84
Percent Similarity: 53.40% Conservative: 26
Best Local Similarity: 40.78% Mismatches: 59
Query Match: 14.01% Indels: 38
DB: 2 Gaps: 5
US-09-902-772-2 (1-451) x US-08-469-412A-1 (1-2667)
QY 239 ProThrGlnSerLysAlaThrGlnProSerSerThrValProLysThrGluAsp--G 258
Db 74 CCGGGAGCCGGCCGGGATCGGGCGCTTCGCCCGGCCCGCCACCATGAAGACCC 133
QY 258 InArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsp 278
Db 134 GCGGACACAGAGGTTGCTTCCTCCCGATT---GGCCCTACA---AGCCAGAG-TGCTCCC 186
QY 278 rGlySerGlyGlnIleGlnLeuThrGlnPheLeuLeuGlnLeuSerSerSer 298
Db 187 CTGGCTCAGGAGATCCAGCTGTGCACATTATCTTGAGCTGCTGCGGAGGAGGAGT 246
QY 298 snSerAsnCysIleThrTyrGluGlyThrAsnGlyGluPheLysMetThrAspProAspG 318
Db 247 ACCAGGGCGTTCATTGCTGGCAGGGGAGTACGGGGGAATTGTCATCAAGACCCCTGATG 306
QY 318 luValAlaArgArgTyrGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuS 338
Db 307 AGGTGGCCCGGCTGTGGGGCGTTCGCAAGTGCAGCCCGCAGATGAATTACGACAAGCTGA 366
QY 338 erArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgT 358
Db 367 GCGGGGCGCTGCGCTATTACTATACAAAGCGCATTCGCACAAAGCAAGGAGGAAACGCT 426
QY 358 yrAlaTyrLysPheAspPheHis----- 365
Db 427 TCACCTACAAGTTCAATTTCAACAAACGGTGGTGGTCAATTACCATTCATTGATGTGG 486
QY 366 -----GlyIleAlaGlnAlaLeuGlnProHisProGluProGluSerSerMetT 381
Db 487 GGTGGTGGGGTGCAGTGGCCCGCAGAGTGGCCCGCAGTGGCGTGGGTGGTAGCCACT 546
QY 381 yrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisProGlnLysMetA 401
Db 547 TCGCTTCCTCCCTCAAGCCCC-----TCCG 573
QY 401 snPheValAlaPro-----HisProAlaLeuProValThrSerSerS 416
Db 574 AGGTGCTGTCCTCCCGCAGGAGCCCGCGCTCACACCGAGCTGCTCTTCATCTTCATCTT 633
QY 416 erPhePheAlaAla 420
Db 634 CCTCTTCTCGGCT 647
RESULT 11
US-09-021-715-1
; Sequence 1, Application US/09021715
; Patent No. 6194547
; GENERAL INFORMATION:
; APPLICANT: Mavrothalassitis, George J.
; APPLICANT: Blair, Donald G.
; APPLICANT: Fisher, Robert J.
; APPLICANT: Beal Jr., Gregory J.
; APPLICANT: Athanasios, Meropi A.
; APPLICANT: Sgouras, Dionysios N.
; TITLE OF INVENTION: The ERF Genetic Locus and Its Products
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/021,715
; FILING DATE: 10-Feb-1998
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Garrett-Wackowski, Eugenia
; REGISTRATION NUMBER: 37,330
; REFERENCE/DOCKET NUMBER: 015280-229000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2667 base pairs
; TYPE: nucleic acid

```

```

1 TITLE OF INVENTION:  CANCER CONTROL
2
3 NUMBER OF SEQUENCES:  5
4 CORRESPONDENCE ADDRESS:
5 ADDRESSEE:  BROWDY AND NEIMARK, P.L.L.C.
6 STREET:  624 Ninth Street N.W., Ste. 300
7 CITY:  Washington
8 STATE:  D.C.
9 COUNTRY:  USA
10 ZIP:  20001
11
12 COMPUTER READABLE FORM:
13 MEDIUM TYPE:  Floppy disk
14 COMPUTER:  IBM PC compatible
15 OPERATING SYSTEM:  PC-DOS/MS-DOS
16 SOFTWARE:  Patent In Release #1.0, Version #1.30
17 CURRENT APPLICATION DATA:
18 APPLICATION NUMBER:  US/08/875,944B
19 FILING DATE:  07-AUG-1997
20 CLASSIFICATION:  514
21 PRIOR APPLICATION DATA:
22 APPLICATION NUMBER:  JP 07-020173
23 FILING DATE:  08-FEB-1995
24 PRIOR APPLICATION DATA:
25 APPLICATION NUMBER:  PCT/JP96/00016
26 FILING DATE:  09-JAN-1996
27 ATTORNEY/AGENT INFORMATION:
28 NAME:  BROWDY, Roger L.
29 REGISTRATION NUMBER:  25,618
30 REFERENCE/DOCKET NUMBER:  FUJINAGA=1
31 TELECOMMUNICATION INFORMATION:
32 TELEPHONE:  (202) 737-3528
33 TELEFAX:  (202) 737-3528
34
35 INFORMATION FOR SEQ ID NO: 1:
36 SEQUENCE CHARACTERISTICS:
37 LENGTH:  2064 base pairs
38 TYPE:  nucleic acid
39 STRANDEDNESS:  single
40 TOPOLOGY:  linear
41 MOLECULE TYPE:  cDNA
42 FEATURE:
43 NAME/KEY:  CDS
44 LOCATION:  1..1386
45
46 US-08-875-944B-1
47
48 Alignment Scores:
49
50 Pred. No.:  5,62e-24      Length:  2064
51 Score:  322.00      Matches:  129
52 Percent Similarity:  38.41%      Conservative:  50
53 Best Local Similarity:  27.68%      Mismatches:  152
54 Query Match:  13.18%      Indels:  137
55 DB:  3      Gaps:  17
56
57 US-09-902-772-2 (1-451) x US-08-875-944B-1 (1-2064)
58
59 QY  45 SerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnProAlaArgVal 64
60 DB  218 TCAAGAGGAGGCCCGACAGTCTCCGCACAG----- 247
61
62 QY  65 ThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSerProAsp 84
63 DB  248 -----ACCGGCCCTGTCTCTGCAGCAGGAGCGCCACTCCCTTACC--- 289
64
65 QY  85 CysSerValAlaLysGlyLysMetValSerSerSerAspAsnValGlyMetAsnTyr 104
66 DB  290 -----ACCATGGCGCAGCAGTGCC----- 307
67
68 QY  105 GlySerTyrMetGluGluLysHisIleProProAsnMetThrThrAsnGluArg 124
69 DB  308 -----TTTACTCCAGTGCCTATGACCCCCCAGACAATCGCCATCAAGTCCCTG 358
70
71 QY  125 ValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGlnTrpLeu 144
72 DB  359 CCCCCTGGTCCCTTGGACAGT-----CGCCCCCTACAGCCCT 394

```

145 TrpAlaValIysGluTyrGly-----LeuProAspValAspIleLeuLeu 159  
: : : : :  
395 TTCCCGGGCAGACGAAATTCCTGAGATCCTCTGGCACTCCCGCCACCCCTG 454  
160 PheGlnAsnIleAspGlyLeuGluCysLysMetThrLysAspPheGlnArgLeu 179  
: : : : :  
455 GCCATGGGTACTCGGG-----AACATA  
180 ThrProSerTyrAsnAlaAspIleLeuSerHisLeuHisTyrLeuArgGluArgGly 199  
: : : : :  
479 GCTCGGTCTCCAGAGCCCTGACATTCCTGCACTCTTCACATCTCAGGAGGGGCC 538  
200 AlaThrPheIlePhePro-----AsnThrSer-----ValTyrProGluAla 213  
: : : : :  
539 GGGAAACCCCTCCAGCCCTTACCAACACCACTGTGCGAGCCCTGCCACCTATCCCC 598  
214 ThrGlnArgIleThrThrArgProAspLeuProTyrGluGlnAlaArgAspSer----- 231  
: : : : :  
599 AGCAGAGCTTTAAGCAAGAATACCATGATCCCTGTATGAACAGCGCGGCCAGCCG 658  
232 -----AlaThrThrSerHisSer 237  
659 TGGACCAAGGTGGGTCAATGGGCACAGGTACCCAGGGCGGGGTGTGATCAAAACAGG 718  
238 HisProThrGlnSerLysAlaThrGlnProSerSerThrValProLysThrGluAsp 257  
: : : : :  
719 AACAGACGACTTCGCCTACACTCAGATGCACGGGTGGCGATCAATGTACTCCACA 778  
258 GlnArgProGlnLeuAspProTyrGlnIleLeuGlyPro----- 270  
: : : : :  
779 CAGAGGGTCTCTGGGCCCTCTCCAGGTGACGGGCCATGGGTATGGCTATGAGAAC 838  
271 -----ThrSer 272  
839 CTCTGCGACCATCCAGATGATGTCTGCTGTCTCCGTGAGAAATTTGAAGGAGACATCA 898  
273 SerArgLeuAlaAsnProGly-----SerGlyGlnIle 283  
: : : : :  
899 AGCAGGAAGGGTGGTGCATTCGAGAGGGCGGCCCTACAGCGCGGGGTGCCCTG 957  
284 GlnLeuThrGlnPheLeuLeuGluLeuSerAspSerSerAsnSerAsnCysIleThr 303  
: : : : :  
958 CAGCTGTGGCAATTTCTGGTGGCTCTGCTGATGACCCACCAAAATGCCCATTTCAITGCC 1017  
304 TrpGluGlyThrAsnGlyGluPheLysMetThrAspProAspGluValAlaArgTyr 323  
: : : : :  
1018 TGGACGGCCGGGATCGAGTTCAAGCTCATTGAGCTGAGGAGTGCAGGCTCTGG 1077  
324 GlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerAlaLeuArgTyr 343  
: : : : :  
1078 GGCATCCAGAAGACCGGCCAGCATGAATACGACAACTGAGCGCTGCTCCGATAC 1137  
344 TyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaTyrLysPheAsp 363  
: : : : :  
1138 TATTATGAGAAGGCATCATGCAAGGTGGCTGGTGAAGCTTACGTGTACAGTTT--- 1194  
364 PheHisGlyIleAlaGlnAla-LeuGlnProHisProGluSerSerMetTyrLysTyr 383  
: : : : :  
1195 -----GTGTGTGAGCCGAGGCC-----TCTTCTCT---TG 1224  
383 rProSer-----AspLeuProTyrMetSerSerTyrHisAlaHisProGln 398  
: : : : :  
1225 GCCTTCCCGCAATACAGCTCCAGCTCTCAAGGTGAGTTTGACCGGCTGTGTCAGTGAG 1284  
398 nLysMetAsnPheValAlaProHisProAlaLeuProValThrSerSerSerPheP 418  
: : : : :  
1285 GAGGACAGCTCCCTTTGCTCCACTTGGATGAGACCCCGCTACTCCAGAGCTGGCT 1344  
418 eAlaAlaProAsnProTyrTrpAsnSerProThrGlyGlyIleTyr-----ProAsnTh 436  
: : : : :  
1345 GSCCCGCGCCAG-CCATTT-----GGCCCCAAGGGTGGCTACTCTTACTAGCCCCCAGC 1397  
436 rArgLeuProAlaAla 441

1398 GGCTGTTCCTCCCTGGC 1413  
: : : : :  
RESULT 13  
US-09-116-049-3  
; Sequence 3, Application US/09116049A  
; Patent No. 6248351  
; GENERAL INFORMATION:  
; APPLICANT: Hung, Men-Chie  
; TITLE OF INVENTION: HUMAN PEA3 IS A TUMOR SUPPRESSOR FOR CANCER CELLS  
; FILE REFERENCE: UTSC:592  
; CURRENT APPLICATION NUMBER: US/09/116, 049A  
; CURRENT FILING DATE: 1998-07-15  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 2064  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-116-049-3  
Alignment Scores: Length: 2064  
Pred. No.: 5,62e-24 Matches: 129  
Score: 322.00 Conservative: 50  
Percent Similarity: 38.41% Mismatches: 152  
Best Local Similarity: 27.68% Indels: 137  
Query Match: 13.18% Gaps: 17  
DB: 17  
US-09-902-772-2 (1-451) x US-09-116-049-3 (1-2064)  
QY 45 SerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnProProAlaArgVal 64  
: : : : :  
DB 218 TCAGAGAGAGGCCAGAGTCCCGCACAG----- 247  
QY 65 ThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSerProAspAsp 84  
: : : : :  
DB 248 -----ACCGGGCCCTGTCTCTGACGAGGAGGAGCGCCACCTCCCTATCC 289  
QY 85 CysSerValAlaLysGlyGlyLysMetValSerSerSerAspAsnValGlyMetAsnTyr 104  
: : : : :  
DB 290 -----ACCATGGGACAGTGGC----- 307  
QY 105 GlySerTyrMetGluGluLysHisIleProProProAsnMetThrThrAsnGluArgArg 124  
: : : : :  
DB 308 -----TTTACTCCAGTGGCTATGACCCCGCCAGACAAATCGCCATCAAGTCCCTG 358  
QY 125 ValIleValProAlaAspProThrLeuTyrSerThrAspHisValArgGlnTrpLeuGlu 144  
: : : : :  
DB 359 CCCTCGTGGTCCCTTGGACAGT-----CGCCCTACAGCCCT 394  
QY 145 TrpAlaValIysGluTyrGly-----LeuProAspValAspIleLeuLeu 159  
: : : : :  
DB 395 TTCCCGGGCAGACGAAATTCCTGAGATCCTCTGGCACTCCCGCCACCCCTG 454  
QY 160 PheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspPheGlnArgLeu 179  
: : : : :  
DB 455 GCCATGGGTACTCGGG-----AACATA 478  
QY 180 ThrProSerTyrAsnAlaAspIleLeuSerHisLeuHisTyrLeuArgGluArgGly 199  
: : : : :  
DB 479 GCTCGGTCTCCAGAGCCCTGACATTCCTGCACTCTTCACATCTCAGGAGGGGCC 538  
QY 200 AlaThrPheIlePhePro-----AsnThrSer-----ValTyrProGluAla 213  
: : : : :  
DB 539 GGGAAACCCCTCCAGCCCTTACCAACACCACTGTGCGAGCCCTGCCACCTATCCCC 598  
QY 214 ThrGlnArgIleThrThrArgProAspLeuProTyrGluGlnAlaArgAspSer----- 231  
: : : : :  
DB 599 AGCAGAGCTTTAAGCAAGAATACCATGATCCCTGTATGAACAGCGCGGCCAGCCG 658  
232 -----AlaThrThrSerHisSer 237  
: : : : :  
QY



```
QY 200 AlaThrPheIlePhePro-----AsnThrSer-----ValTyrProGluAla 213
Db 539 GGGAAACCCCTCCAGCCCTTACCAACACACAGCTGTGGAGCCCTGCCACCTTATCCCC 598
QY 214 ThrGlnArgIleThrThrArgProAspLeuProTyrGluGlnAlaArgArgSer----- 231
Db 599 AGCAGAGCTTTAAGCAAGAAATACATGATCCCTGTATGAACAGCGCGGCCAGCCGCG 658
QY 232 -----AlaThrPheIlePhePro-----AlaThrPheIlePhePro----- 231
Db 659 TGGACACAGGCTGGGTCAATGGGCACAGTACCCAGCGCGGGGGTGTGATCAACAGG 718
QY 238 HisProThrGlnSerLysAlaThrGlnProSerSerThrValProLysThrGluAsp 257
Db 719 AACAGACGAGCTTCGCTACGACTCAGATGTCACCGGTGGCATCAATGTACCTCCACA 778
QY 258 GlnArgProGlnLeuAspProTyrGlnIleLeuGlyPro----- 270
Db 779 CAGAGGGCTTCTCTGGGCCCTCTCCAGGTGACGGGGCCATGGGCTATGAGTATGAGAAAC 838
QY 271 -----ThrSer 272
Db 839 CTCTCGACCATTCACAGATGATGCTGCTGTTGCTCCCTGAGAAATTTGAAGGACATCA 898
QY 273 SerArgLeuAlaAsnProGly-----SerglyGlnIle 283
Db 899 AGCAGAGGGGTGGT-GCATTTCGAGAGGGGGCCCTACCAGCGCGGGGTGCGCTG 957
QY 284 GlnLeuThrGlnPheLeuLeuLeuLeuSerAspSerSerSerSerSerSerSerSerSer 303
Db 958 CAGCTGTGGCAATTTCTGGTGGCTTCTGGATGATGATGATGATGATGATGATGATGATG 1017
QY 304 TrpGluGlyThrAsnGlyGluPheLysMetThrAspProAspGluValAlaAlaArgArgTyr 323
Db 1018 TGGACGGCGCGGGAATGGAGTCAAGCTCATGAGCTGAGGAGTGGCGAGGCTCTGG 1077
QY 324 GlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArgAlaLeuArgTyr 343
Db 1078 GGCATCCAGAAAGGCGCCAGCCAGCCAGTGAATACGACAGCTGAGCGCTGCTCCGATAC 1137
QY 344 TyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaTyrLysPheAsp 363
Db 1138 TATTATGAGAAAGGCGATCATCAGAGGTGGCTGGTGGAGCTTACGTTACAGTTT--- 1194
QY 364 PheHisGlyIleAlaGlnAla-LeuGlnProHisProProGluSerSerMetTyrLysTyr 383
Db 1195 -----GTGTGTGAGCCCGAGGCC-----TCTTCTCTT-----TG 1224
QY 383 rProSer-----AspLeuProTyrMetSerSerTyrHisAlaHisProGly 398
Db 1225 GCCTTCCGGACAATCAGCGTCCAGCTCTCAGGCTGAGTTGACCGGCTGTGAGTGAG 1284
QY 398 nLysMetAsnPheValAlaProHisProProAlaLeuProValThrSerSerSerPheP 418
Db 1285 GAGGACAGAGTCCCTTTGCTCCACTTGGATGAGAGCGCGGCTACCTCCAGAGCTGGCT 1344
QY 418 eAlaAlaProAsnProTyrTyrAsnSerProThrGlyGlyIleTyr-----ProAsnTh 436
Db 1345 GCGCCCGCCAG-CCATTT-----GGCCCCAAGGGTGGCTACTCTTACTAGCCCCCAGC 1397
QY 436 rArgLeuProAlaAla 441
Db 1398 GGCTGTTCCTCCCTGCC 1413
```

## RESULT 15

```
US-09-884-363-3
; Sequence 3, Application US/09884363
; Patent No. 6582725
; GENERAL INFORMATION:
; APPLICANT: Hung, Men-Chie
; TITLE OF INVENTION: HUMAN PEA3 IS A TUMOR SUPPRESSOR FOR CANCER CELLS
; FILE REFERENCE: UTSC:562
```

```
; CURRENT APPLICATION NUMBER: US/09/884,363
; CURRENT FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 09/116,049
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2064
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-884-363-3

Alignment Scores:
Pred. No.: 5,62e-24 Length: 2064
Score: 322.00 Matches: 129
Percent Similarity: 38.41% Conservative: 50
Best Local Similarity: 27.68% Mismatches: 152
Query Match: 13.18% Indels: 137
DB: Gaps: 17

US-09-902-772-2 (1-451) x US-09-884-363-3 (1-2064)

QY 45 SerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnProProAlaArgVal 64
Db 218 TCARGAGAGGCCAGAGTCCCGCAGCAG----- 247
QY 65 ThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSerProAspAsp 84
Db 248 -----ACCGCGCCCTGCTCTGCAGCAGGAGCGCCACTCCCTACCC--- 289
QY 85 CysSerValAlaLysGlyGlyLysMetValSerSerSerSerSerSerSerSerSer 104
Db 290 -----ACCATGGCGAGAGTCCC----- 307
QY 105 GlySerTyrMetGluGluLysHisIleProProProAsnMetThrThrAsnGluArgArg 124
Db 308 -----TTTACTCCAGTGCCTATGACCCCGCCAGCAAAATCGCATCAAGTCCCTG 358
QY 125 ValIleValProAlaAspProThrLeuTyrSerThrAspHisValArgGlnTrpLeuGlu 144
Db 359 CCCTGCTGCTCCCTTGACAGT-----CGCCCCATCAGCCCT 394
QY 145 TrpAlaValLysGluTyrGly-----LeuProAspValAspIleLeuLeu 159
Db 395 TTCGCCGCGCAGAGCAACGGAATTTCTGAGATCTCTGGCACCTCCCGACCCACCCCTG 454
QY 160 PheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeu 179
Db 455 GCATGGGTACCTCGGG-----AACATA 478
QY 180 ThrProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArgGly 199
Db 479 GCTCCGTCTTCCAGCAGCCCTGGACATTTGCACATCTTCATCTCAGGAGGGGGCC 538
QY 200 AlaThrPheIlePhePro-----AsnThrSer-----ValTyrProGluAla 213
Db 539 GGGAAACCCCTCCAGCCCTTACCAACACAGCTGTGGAGCCCTGCCACCTTATCCCC 598
QY 214 ThrGlnArgIleThrThrArgProAspLeuProTyrGluGlnAlaArgArgSer----- 231
Db 599 AGCAGAGCTTTAAGCAAGAAATACATGATCCCTGTATGAACAGCGCGGCCAGCCGCG 658
QY 232 -----AlaThrPheIlePhePro-----AlaThrPheIlePhePro----- 237
Db 659 TGGACACAGGCTGGGTCAATGGGCACAGTACCCAGCGCGGGGGTGTGATCAACAGG 718
QY 238 HisProThrGlnSerLysAlaThrGlnProSerSerThrValProLysThrGluAsp 257
Db 719 AACAGACGAGCTTCGCTACGACTCAGATGTCACCGGTGGCATCAATGTACCTCCACA 778
QY 258 GlnArgProGlnLeuAspProTyrGlnIleLeuGlyPro----- 270
Db 779 CAGAGGGCTTCTCTGGGCCCTCTCCAGGTGACGGGGCCATGGGCTATGAGTATGAGAAAC 838
```

```
QY 271 -----ThrSer 272
Db 839 CTCTGCGACCATCCCGATGATCTCGGTGTCCTCGAGAAATTTGAAGGAGACATCA 898
QY 273 SerArgLeuAlaAsnProGly-----SerGlyGlnIle 283
Db 899 AGCAGGAAGGGTCCGT-GCATTTCGAGAGGGGCGCCCTACCAGCGCGGGGTGCCCTG 957
QY 284 GlnLeuTrpGlnPheLeuLeuLeuSerSerSerSerSerSerSerSerSerSerSer 303
Db 958 CAGCTGTGGCAATTTCTGGTGGCTTGTGATGACCCAAACAAATGCCCATTTCAATGCC 1017
QY 304 TrpGluGlyThrAsnGlyGluPheLysMetThrAspProAspGluValAlaArgArgTrp 323
Db 1018 TGGACGGCGCGGGAATGGAGTTCAAGCTCATTGAGCTGAGGAGGTCCGAGGCTCTGG 1077
QY 324 GlyGluArgLysSerLysProAsnMetAsnTrpAspLysLysLeuSerArgAlaLeuArgTyr 343
Db 1078 GGCATCCAGAAAGACCGCCAGCCATGAATTACGACAAAGCTGAGCGCTCGCTCCGATAC 1137
QY 344 TyrTrpAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaTyrLysPheAsp 363
Db 1138 TATTATGAAAGGCAATCATCAGAGGTGCTGCTGAGCGTTACGTGTACAGTTT--- 1194
QY 364 PheHisGlyIleAlaGlnAla-LeuGlnProHisProHisProGluSerSerMetTyrLysTyr 383
Db 1195 -----GTGTGTGAGCCCGAGGCC-----TCTTCTCTT-----TG 1224
QY 383 rProSer-----AspLeuProTyrMetSerSerTyrHisAlaHisProGlu 398
Db 1225 GCCTTCCCGGACATCAGCGTCCAGCTCTCAGGCTGAGTTGACCGGCTGTCCAGTGAG 1284
QY 398 nLysMetAsnPheValAlaProHisProProAlaLeuProValThrSerSerSerPhePh 418
Db 1285 GAGGACACAGTCCCTTTGTCCCACTTGGATGAGAGCCCGCCTACCTCCAGAGCTGGCT 1344
QY 418 eAlaAlaProAsnProTyrTrpAsnSerProThrGlyGlyIleTyr-----ProAsnTh 436
Db 1345 GGCCCCGGCCAG-CAATTT-----GGCCCCAAGGGTGGGTACTCTTACTAGCCCCCAGC 1397
QY 436 rArgLeuProAlaAla 441
Db 1398 GGCTGTTCCCTCCCTGCC 1413
```

Search completed: May 27, 2004, 23:58:56  
Job time : 149 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 27, 2004, 22:55:58 ; Search time 668 Seconds  
(without alignments)  
3071.189 Million cell updates/sec

Title: US-09-902-772-2

Perfect score: 2444

Sequence: 1 MASTIKKALSVSDQSLPE.....IYENTRLPRAHMPHSLGTYV 451

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2960401 seqs, 2274450654 residues

Total number of hits satisfying chosen parameters: 5920802

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US0902772/runat\_27052004\_162843\_2259/app\_query.fasta\_1.647  
-DB\_Published Applications NA -OPT=fastp -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cd1 -LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US0902772@cgn\_1\_1\_481@runat\_27052004\_162843\_2259  
-NCPU=6 -ICPU=3 -NO\_MMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq2:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	2444	100.0	1447	9	US-09-902-772-1
2	2414.5	98.8	1528	9	US-09-902-772-3
3	2182	89.3	1389	13	US-10-211-462-41
4	2182	89.3	1389	13	US-10-087-192-1127
5	2182	89.3	1366	15	US-10-205-823-98
6	2182	89.3	1366	15	US-10-021-660-26
7	2057	84.2	1297	13	US-10-087-192-1124
8	1624.5	66.5	1359	12	US-10-052-482-198
9	1624.5	66.5	2957	13	US-10-052-482-197
10	1624.5	66.5	2957	13	US-10-342-887-710
11	1624.5	66.5	2957	13	US-10-172-118-710
12	1624.5	66.5	2957	13	US-10-007-926A-295
13	1613.5	66.0	1359	12	US-10-052-482-195
14	1613.5	66.0	1729	12	US-10-052-482-194
15	956.5	39.1	21231	13	US-10-087-192-1126
16	949	38.8	69090	13	US-10-087-192-1123
17	948	38.8	567	9	US-09-864-761-18410
18	863	35.3	473	9	US-09-864-761-20472
19	857	35.1	472	15	US-10-029-386-25224
20	833.5	34.1	73725	12	US-10-052-482-196
21	813.5	33.3	72732	12	US-10-052-482-193
22	712	29.1	549	9	US-09-923-779-54
23	650	26.6	420	9	US-09-864-761-3705
24	650	26.6	454	9	US-09-864-761-1853
25	638	26.1	467	10	US-09-918-995-23356
26	511	20.9	1752	9	US-09-850-799-1
27	511	20.9	1752	15	US-10-037-859-1
28	453.5	18.6	380	15	US-10-263-828-41
29	451	18.5	1496	13	US-10-087-192-344
30	449	18.4	1450	16	US-10-151-997-126
31	449	18.4	1602	13	US-10-087-192-347
32	446	18.2	4991	12	US-10-152-319A-1502
33	445.5	18.2	2269	9	US-09-954-531-955
34	444.5	18.2	2188	9	US-09-920-300A-1716
35	444.5	18.2	2188	14	US-10-033-528-1716
36	444.5	18.2	2188	15	US-10-039-926-1716
37	444.5	18.2	2268	9	US-09-920-300A-1693
38	444.5	18.2	2268	14	US-10-033-528-1693
39	444.5	18.2	2268	15	US-10-099-926-1693
40	444.5	18.2	3692	15	US-10-106-698-824
41	441.5	18.1	2777	16	US-10-108-260A-1528
42	431.5	17.7	1884	9	US-09-925-300-420
43	407	16.7	2532	15	US-10-037-270-869
44	407	16.7	2532	16	US-10-117-722-869
45	401.5	16.4	3695	13	US-09-805-020-17

Alignment Scores:  
Pred. No.: 3.57e-269 Length: 1447

#### ALIGNMENTS

RESULT 1  
US-09-902-772-1  
; Sequence 1, Application US/09902772  
; Patent No. US20020164739A1  
; GENERAL INFORMATION:  
; APPLICANT: Chugai Seiyaku Kabushiki Kaisha, 5001, Iwamoto et  
; TITLE OF INVENTION: Cell Calcification Suppressing Proteins and Genes of  
; TITLE OF INVENTION: the Proteins  
; FILE REFERENCE: chugai seiyaku kabushiki kaisha 5001  
; CURRENT APPLICATION NUMBER: US/09/902,772  
; CURRENT FILING DATE: 2001-07-12  
; PRIOR APPLICATION NUMBER: US/08/878,177  
; PRIOR FILING DATE: 1997-06-18  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO: 1  
; LENGTH: 1447  
; TYPE: DNA  
; ORGANISM: C-11 gene, c-erg gene w/ deletion, chicken DNA  
US-09-902-772-1



```
Score: 2444.00 Matches: 451
Percent Similarity: 100.00% Conservativity: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-902-772-2 (1-451) x US-09-902-772-1 (1-1447)
QY 1 MetAlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20
DB 63 ATGGCAGACACTATTAGGAAGCATTTATCAGTGGTGAAGCAAGCAGCTCTGTTGAG 122
QY 21 CysAlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGlu 40
DB 123 TGTGCTACGGATCGCCCACTTTCGAAAGACAGAAATGACAGCTCTCTCCAGTGAA 182
QY 41 TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnPro 60
DB 183 TATGGCAACATCAAGATGAGCCCGCGTTCACCAAGCAGGACTGGTTATACAGCCC 242
QY 61 ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
DB 243 CCGGCCAGAGTTACATTAGATGGAGTGAACCCAAACAGGTTAATGGGTCAAGGAAT 302
QY 81 SerProAspPheCysSerValAlaLysGlyGlyLeuMetValSerSerSerAspAsnVal 100
DB 303 TCACCTGATGACTCGACGCGTGGCAAAAGGAGGAAATGTTAGCAGTTTCAAGCAATGTT 362
QY 101 GlyMetAsnTyrGlySerTyrMetGluGluLysHisLeuProProAsnMetThrThr 120
DB 363 GGGATGAACATATGGAAGCTACATGGAAGAGAGCATATTCCGCTCCAAATATGACAACC 422
QY 121 AsnGluArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArg 140
DB 423 AATGAACGAGAGTATTGTGCCAGCAGATCTTACGTATGGAGCAGACCACTATGACGG 482
QY 141 GlnTrpLeuGlnTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPhe 160
DB 483 CAGTGGCTGGAGTGGCAGTGAAGAGTATGTTCTTCCAGAGCTGGACATCTTGTGTTTC 542
QY 161 GlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspPheGlnArgLeuThr 180
DB 543 CAGAACATTGATGGGAAGAGTGTGTGTAATGATCCAAAGATGACTTCCAGAGACTCAGC 602
QY 181 ProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArgGlyAla 200
DB 603 CCGAGCTATAACGAGATATCTCTCTGTCACACTACCTACCTCAGAGAGAGAGGCC 662
QY 201 ThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrArg 220
DB 663 ACTTTTATTTTCCAAATATACATCAGTTTACCCAGAGCAACGCAAGAAATAACAACAAGG 722
QY 221 ProAspLeuProTyrGluGlnAlaArgSerAlaTrpThrSerHisSerHisProThr 240
DB 723 CCAGATTACCTTATGAGCAAGCAGAGAGATCAGCGTGGAGAGTCAACCCATCCCACT 782
QY 241 GlnSerLysAlaThrGlnProSerSerThrValProLysThrGluAspGlnArgPro 260
DB 783 CAGTCAAAAGCTACCAACCATCATCTTCAACAGTGCCCAAAACAGAAAGCAGCGTCT 842
QY 261 GlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGlySer 280
DB 843 CAGTTAGATCTTATCAGATTCTTGGACCGCAGCAGCGCTCTTGGAAATCCAGGAGT 902
QY 281 GlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeuSerAspSerSerAsnSerAsn 300
DB 903 GGGCAGATACAGCTATGCGAGTCTCTCTGAGCTTCTGTCGAGAGCTCCAACCTCCAAC 962
QY 301 CysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspProAspGluValAla 320
DB 963 TGCATCACTGGGAGGGCACAATAATGGGAGTCAAGATGACAGACCTGATGAAGTGCT 1022
QY 321 ArgArgTrpGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArgAla 340
```

```
1023 CGGCGTTGGGAGAGAGCAAAAGCAACCTAACATGAATGACAAACTCAGCGGTGCA 1082
QY 341 LeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaTyr 360
DB 1083 CTTGCTACTACTATGACAAATAATATTGACTAAAGTTTATGTTAAACGCTATGCTTAC 1142
QY 361 LysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGluSerSerMet 380
DB 1143 AAATTTGATTTCCACGGAATCGCTCAGGCCCTCCAGGCTCACCTCCAGAAATCATCCATG 1202
QY 381 TyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisProGlnLysMet 400
DB 1203 TACAAATACCATCAGACCTCCCTTACATGAGTTCTCCATCCAGCACCCCAAGATG 1262
QY 401 AsnPheValAlaProHisProAlaLeuProValThrSerSerSerPhePheAlaAla 420
DB 1263 AACTTTGTAGCTCCCAATCCCTGCTTTGCGCGTAACCTCATCCAGCTTTTGTGCTGCC 1322
QY 421 ProAsnProTyrTrpAsnSerProThrGlyGlyLeuTyrProAsnThrArgLeuProAla 440
DB 1323 CCTAATCCATATCGGAATTCACCAACTGGAGGCACTACCCCAATACCCAGGCTGCCAGCT 1382
QY 441 AlaHisMetProSerHisLeuGlyThrTyrTyr 451
DB 1383 GCTCATATGCTTCCCATCTTGGCACCTACTAC 1415

RESULT 2
US-09-902-772-3
; Sequence 3, Application US/09902772
; Patent No. US20020164739A1
; GENERAL INFORMATION:
; APPLICANT: Chugai Seiyaku Kabushiki Kaisha, 5001, Iwamoto et
; TITLE OF INVENTION: Cell Calcification Suppressing Proteins and Genes of
; TITLE OF INVENTION: the Proteins
; FILE REFERENCE: chugai seiyaku kabushiki kaisha 5001
; CURRENT APPLICATION NUMBER: US/09/902,772
; PRIOR FILING DATE: 2001-07-12
; CURRENT APPLICATION NUMBER: US/08/878,177
; PRIOR FILING DATE: 1997-06-18
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1528
; TYPE: DNA
; ORGANISM: c-erg gene, chicken DNA
US-09-902-772-3

Alignment Scores:
Pred. No.: 9 186-286 Length: 1528
Score: 2414.50 Matches: 450
Percent Similarity: 94.14% Conservativity: 0
Best Local Similarity: 94.14% Mismatches: 1
Query Match: 98.79% Indels: 27
DB: 9 Gaps: 1

US-09-902-772-2 (1-451) x US-09-902-772-3 (1-1528)
QY 1 MetAlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20
DB 63 ATGGCAGACACTATTAGGAAGCATTTATCAGTGGTGAAGCAAGCAGCTCTCTGTTGAG 122
QY 21 CysAlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGlu 40
DB 123 TGTGCTACGGATCGCCCACTTTCGAAAGACAGAAATGACAGCTCTCTTCCAGTGAA 182
QY 41 TyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnPro 60
DB 183 TATGGCAACATCAAGATGAGCCCGCGTTCACCAAGCAGGACTGGTTATACAGCCC 242
QY 61 ProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsn 80
DB 243 CCGGCCAGAGTTACATTAGATGGAGTGAACCCAAACAGGTTAATGGGTCAAGGAAT 302
```

QY 81 SerProAspCysSerValAlaLysGlyGlyLysMetValSerSerSerAspVal 100  
Db 303 TCACCTGATGACTGACGCGTGGCAAGAGGAGGAAATGGTTAGCAGCTTCAGACATGTT 362  
QY 101 GlyMetAsnTyrGlySerTyrMetGluGluLysHisLeuProProAsnMetThrThr 120  
Db 363 GGGATGAACATATGGAAGCTACATGGAAGAGAAAGCATATTCGGCTCCCAATATGCAAAAC 422  
QY 121 AsnGluArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArg 140  
Db 423 AATGAACGAAGAGTTATTTGCCAGCAGATCTTACGTTNTGGAGCAGACACCATGTACGG 482  
QY 141 GlnTrpLeuGluTrpAlaValLysGlyLysGlyLeuProAspValAspIleLeuLeuPhe 160  
Db 483 CAGTGGCTGGAGTGGCAGTGAAGAGTATGGTCTTCAGACGCTGGACATCTTTGTTGTC 542  
QY 161 GlnAsnIleAspGlyLysGluLeuLysCysLysMetThrLysAspAspPheGlnArgLeuThr 180  
Db 543 CAGAACATTTGATGGGAAGAGTGTGTAAATGACCAAGATGACTTCCAGAGACTCAGC 602  
QY 181 ProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArg----- 198  
Db 603 CCGAGCTATAACGAGATATCTCTGTACACCTACACTACCTCAGAGAGACTCCTCTT 662  
QY 198 ----- 198  
Db 663 CCACATTTGACTTCAGATGATGTTGATAAGGCTTTACAAAACTCTCCACGCTTAATGCAT 722  
QY 199 -----GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAla 213  
Db 723 GCTAGAAACACAGAGGAGCCACTTTATTTTCCAAATACATCAGTTTACCAGAGCA 782  
QY 214 ThrGlnArgIleThrArgProAspLeuProTyrGluGlnAlaArgSerAlaTrp 233  
Db 783 ACGCAAGAAATAACAACAGCCAGATTTACCTTATGAGCAAGCGAGAGATCAGCGTGG 842  
QY 234 ThrSerHisSerHisProThrGlnSerLysAlaThrGlnProSerSerSerThrValPro 253  
Db 843 ACGAGTACACCCATCCACATCGTCAAAAGCTACCAACCATCATCTTCAACAGTGCCT 902  
QY 254 LysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSer 273  
Db 903 AAAACAGAGAACCGCGTCTCAGTAGATCTTATCAGATCTTGGCCGACCGCAGCAGC 962  
QY 274 ArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuLeuLeu 293  
Db 963 CGTCTTGCAAAATCCAGGAGTGGGAGATACAGTATGCGCAGTTCTTACTGGAGCTTCTG 1022  
QY 294 SerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMet 313  
Db 1023 TCGNACAGCTCCNACTCCAACTGCATCACCTGGAGGGGCAAAATGGGAGTTCAAGATG 1082  
QY 314 ThrAspProAspGluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMetAsn 333  
Db 1083 ACAGACCTGTATGAATGGCTCGCGTGGGGAGAGAGAGAAAGCAAACTTAAATGAAC 1142  
QY 334 TyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysVal 353  
Db 1143 TATGCAAACTCAGCGGTGCATCTCGCTACTACTATGACAAATAATATGACTTAAAGTT 1202  
QY 354 HisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnPro 373  
Db 1203 CATGGTAAACGCTATGCTTACAAATTTGATTTCCACGGAATCGCTCAGGCGCTCCAGCCT 1262  
QY 374 HisProProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyr 393  
Db 1263 CACCTTCAGATATCATCATGATGACAAATACCATCATCAGACCTTCCCTCATCATGAGTTCTTAC 1322  
QY 394 HisAlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProValThr 413  
Db 1323 CATGCACACCCCCAGAGATGAACCTTTGTAGCTCCCATCCCTGCTTTGCCCGTAAAC 1382

QY 414 SerSerSerPhePheAlaAlaProAsnProTyrTrpAsnSerProThrGlyGlyIleTyr 433  
Db 1383 TCATCCAGCTTTTGTGCTGCCCTAATCCATACTGGAATTCACCAACTGGAGGCATCTAC 1442  
QY 434 ProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451  
Db 1443 CCCAATACAGGCTGCCAGCTCTCATATGCTTCCCATCTTTGGCACCCTACTACTAC 1496  
RESULT 3  
US-10-211-462-41  
; Sequence 41, Application US/10211462  
; Publication No. US20040033495A1  
; GENERAL INFORMATION:  
; APPLICANT: Murray, Richard  
; APPLICANT: Glynn, Richard  
; APPLICANT: Watson, Susan R.  
; APPLICANT: Aziz, Nataasha  
; APPLICANT: Eos Biotechnology, Inc.  
; TITLE OF INVENTION: Methods of Diagnosis of Angiogenesis, Compositions and  
; FILE REFERENCE: 018501-006200US  
; CURRENT APPLICATION NUMBER: US/10/211,462  
; CURRENT FILING DATE: 2003-02-13  
; PRIOR APPLICATION NUMBER: US 09/784,356  
; PRIOR FILING DATE: 2001-02-14  
; PRIOR APPLICATION NUMBER: US 09/791,390  
; PRIOR FILING DATE: 2001-02-22  
; PRIOR APPLICATION NUMBER: US 60/310,025  
; PRIOR FILING DATE: 2001-08-03  
; PRIOR APPLICATION NUMBER: US 60/334,244  
; PRIOR FILING DATE: 2001-11-29  
; NUMBER OF SEQ ID NOS: 230  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 41  
; LENGTH: 1389  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-211-462-41  
Alignment Scores:  
Pred. No.: 3,178-239 Length: 1389  
Score: 2182.00 Matches: 407  
Percent Similarity: 91.72% Conservative: 14  
Best Local Similarity: 88.67% Mismatches: 24  
Query Match: 89.28% Indels: 14  
DB: 13 Gaps: 4  
US-09-902-772-2 (1-451) x US-10-211-462-41 (1-1389)  
QY 2 AlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGluCys 21  
Db 25 GCAGCTCATATCAAGGAAGCTTATCAGTTGTGAGTGAGGACCAAGTCGTTGTTGAGTGT 84  
QY 22 AlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGluTyr 41  
Db 85 GCCTACGGAACCGCACCTCGCTGTAAGCAGAGATGACCGCTCTCTCCAGCGACTAT 144  
QY 42 GlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnProPro 61  
Db 145 GGCACACTTCCAGATGAGCCACCGCTCCCTCAGCAGATTTGGCTGTCTCAACCCCA 204  
QY 62 AlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSer 81  
Db 205 GCCAGGCTCACCATCAAAATGGAATGTAACCTTAGCCAGGTGAATGGCTCAGAGAACTCT 264  
QY 82 ProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsnValGly 101  
Db 265 CTTGATGAATGAGTGTGGCCAAAGCGGGAGAGATGTTGGGAGCCCGCAGACCGTTGG 324  
QY 102 MetAsnTyrGlySerTyrMetGluGluLysHisLeuProProProAsnMetThrThrAsn 121  
Db 325 ATGAATACCGGAGCTACATGAGGAGAGAGCATCCACCCCAACATGACACCAAGAAC 384

122 GluArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGln 141  
Db  
385 GAGCGCAGAGTTATCGTGGCAGCAGATCCCTACGCTATGGAGTACAGACCATGTGGCGCAG 444  
Qy  
142 TrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPheGln 161  
Db  
445 TGGCTGGAGTGGCGGGTGAAGAATATGGCTTCCAGCGTCACATCTTGTATTCCAG 504  
Qy  
162 AsnIleAspGlyLysGluLeuCysLysMetThrLysAspPheGlnArgLeuThrPro 191  
Db  
505 AACATCATGGAGAGAACTGTGCAGATACCCAGAGCAGCTCCAGAGGCTCACCCCC 564  
Qy  
182 SerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArgGlyAlaThr 201  
Db  
565 AGCTACACGGCAGATCTCTCTCATCTCCATCTCCATCTCCAGAGACTCTCTTCCA 624  
Qy  
202 PheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrThrArgPro 221  
Db  
625 CATTTG-----ACTTCA-----GATGATGTTGATAAGCCTTACAAACTCTCCA 669  
Qy  
222 -----AspLeuProTyrGluGlnAlaArgArgSerAlaTrp 233  
Db  
670 CGGTAAATGCATGTAGAAACACAGATTACCATATGAGCCCCCAGAGATCAGCGTG 729  
Qy  
234 ThrSerHisSerHisProThr--GlnSerLysAlaThrGlnProSerSerThrVal 252  
Db  
730 ACCGGTCAGCGCCACCCAGCCCGCTGCAAGAGCTGCTCAACCATCTCTCCACAGTG 789  
Qy  
253 ProLysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSer 272  
Db  
790 CCCAAACTGAAGACACAGCGTCTCAGTTAGATCTTATCAGATCTTGGACCAACAAGT 849  
Qy  
273 SerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeu 292  
Db  
850 AGCGCGCTTGAAATCCAGGCAGTGGCCAGATCCAGCTTTGGCAGTTCTCTCGGAGCTC 909  
Qy  
293 LeuSerAspSerSerAsnSerCysIleThrTyrGluGlyThrAsnGlyGluPheLys 312  
Db  
910 CTGTGGAGACCTCCAACTCCAGTGCATCACCTGGGAAGGCACCAAGGGGAGTTCAG 969  
Qy  
313 MetThrAspProAspGluAlaAlaArgArgTyrGlyGluArgLysSerLysProAsnMet 332  
Db  
970 ATGACGGATCCGACGAGTGGCGCGCTGGGGAGAGCGGAGAGCAAAACCAACATG 1029  
Qy  
333 AsnTyrAspLysLeuSerArgAlaLeuArgTyrTyrTrpAspLysAsnIleMetThrLys 352  
Db  
1030 AACTACGATAAGCTCAGCGCGCCCTCCGTTACTATGACAGAACATCATGACCAAG 1089  
Qy  
353 ValHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGln 372  
Db  
1090 GTCCATGGGAAGCGCTACGCCCTACAAGTTCGACTTCCACGGGATCGCCAGGCCCTCCAG 1149  
Qy  
373 ProHisProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSer 392  
Db  
1150 CCCACCCCGCGGAGTTCATCTCTACAGTACCCCTCAGACCTCCGTCATGGGCTCC 1209  
Qy  
393 TyrHisAlaHisProGlnLysMetAsnPheValAlaProHisProAlaLeuProVal 412  
Db  
1210 TATCACGCCACCCACAGAGATGAATTTGTGGCGCCGCCACCTCCAGCCCTCCCGTG 1269  
Qy  
413 ThrSerSerSerPhePheAlaAlaProAsnProTyrTrpAsnSerProThrGlyIle 432  
Db  
1270 ACATCTTCCAGTTTTTGTGGCGCCCAACCCATCTGGAATTCACCACTGGGGGTATA 1329  
Qy  
433 TyrProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451  
Db  
1330 TACCCCAACACTAGCTCCCCACAGCCATATGCTTCTCATCTGGGCACTTACTAC 1386

## RESULT 4

US-10-087-192-1127  
; Sequence 1127, Application US/10087192  
; Publication No. US20020182586A1  
; GENERAL INFORMATION:

APPLICANT: Morris, David W.  
APPLICANT: Engelhard, Eric K.  
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
TITLE OF INVENTION: CANCER  
FILE REFERENCE: 529452000122  
CURRENT APPLICATION NUMBER: US/10/087,192  
CURRENT FILING DATE: 2002-03-01  
PRIOR APPLICATION NUMBER: US 09/747,377  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: US 09/798,586  
PRIOR FILING DATE: 2001-03-02  
NUMBER OF SEQ ID NOS: 2059  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1127  
LENGTH: 1389  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-087-192-1127  
Alignment Scores:  
Pred. No.: 3,17e-239 Length: 1389  
Score: 2182.00 Matches: 407  
Percent Similarity: 91.72% Conservative: 14  
Best Local Similarity: 88.57% Mismatches: 24  
Query Match: 89.28% Indels: 14  
DB: 13 Gaps: 4  
US-09-902-772-2 (1-451) x US-10-087-192-1127 (1-1389)  
Qy 2 AlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGluCys 21  
Db 25 GCAGCTCATATCAAGAAAGCCCTTATCAGTTGTGAGTGAGGACCACTCGTTGTTGAGTGT 84  
Qy 22 AlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerSerGluTyr 41  
Db 85 GCCTACGGAGACGCCACACCTGGCTAAGACAGAGATGACCGGCTCTCTCTCCAGCGCAT 144  
Qy 42 GlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnProPro 61  
Db 145 GGACAGACTTCCAGATGAGCCCGCTCCCTCAGCAGAGATTGGCTGTCTCAACCCCA 204  
Qy 62 AlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSer 81  
Db 205 GCCAGGGTCACCATCAAAATGGAATGTAACTAGCCAGTGAATGGTCAAGGAACCTCT 264  
Qy 82 ProAspAspCysSerValAlaLysGlyLysMetValSerSerSerSerAspAsnValGly 101  
Db 265 CCGTGAATGAGTGGCGGCGGAGGAGATGGTGGCAGCCCGCCAGACCCGTTGGG 324  
Qy 102 MetAsnTyrGlySerTyrMetGluLysHisIleProProProAsnMetThrThrAsn 121  
Db 325 ATGAACCTACGGCAGCTACATGGAGGAGAAGCACATGCCACCCCAACATGACCAACGAA 384  
Qy 122 GluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGln 141  
Db 385 GAGCCAGAGTTATCGTCCAGCAGATCCTACGCTATGGAGTACAGACCATGTGGCGCAG 444  
Qy 142 TrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPheGln 161  
Db 445 TGGCTGGAGTGGCGGGTGAAGAATATGGCTTCCAGACGCTCAACATCTTGTATTCCAG 504  
Qy 162 AsnIleAspGlyLysGluLeuCysLysMetThrLysAspPheGlnArgLeuThrPro 181  
Db 505 AACATCGATGGGAAGAACTGTGCAAGATGACCAAGGAGCAGCTTCCAGAGGCTCACCCCC 564  
Qy 182 SerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArgGlyAlaThr 201  
Db 565 AGCTACACGGCAGCATCTCTTCTCTCATCTCCATCTCCATCTCCAGAGACTCTCTTCCA 624  
Qy 202 PheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrThrArgPro 221  
Db 625 CATTTG-----ACTTCA-----GATGATGTTGATAAGCCTTACAAACTCTCCA 669

```
QY 222 -----AspLeuProTyrGluGlnAlaArgSerAlaTTP 233
Db 670 CGGTTAATGATGCTAGAAACACAGATTTACCATATGAGCCGCCCCAGGAGATCAGCGCTGG 729
QY 234 ThrSerHisSerHisProThr---GlnSerLysAlaThrGlnProSerSerThrVal 252
Db 730 ACCGGTCACGGCCACCCACAGCCCGCTCGAAGCTGCTCAACATCTCTCTCCACAGTG 789
QY 253 ProLysThrGluAspGlnArgProGlnLeuAspProTyrGlnLeuLeuGlyProThrSer 272
Db 790 CCAAAACTGAAGACAGCGCTCTCAGTTAGATCTTATCAGATTTCTTGACCAACAAGT 849
QY 273 SerArgLeuAlaAenProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeu 292
Db 850 AGCGCCCTTCGAATATCCAGCGAGTGGCCAGATCCAGCTTTGGCAGTTCTCTCTGGAGCTC 909
QY 293 LeuSerAspSerAenSerAenCysIleThrTrpGluGlyThrAsnGlyGluPheLys 312
Db 910 CTGTGGACAGCTCCAACTCCAGCTGCATCACCTGGGAGGCCAACACGGGGAGTTCAAG 969
QY 313 MetThrAspProAspGluAlaAlaArgTyrGlyGluArgLysSerLysProAenMet 332
Db 970 ATGACGGATCCCGACGAGGTGGCCGCGCTGGGGAGCGGAAGACCAACCAACATG 1029
QY 333 AsnTyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLys 352
Db 1030 AACTACGATAAGCTCAGCCGCGCCCTCCGTTACTACTATGACAAGAACATCATGACCAAG 1089
QY 353 ValHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGln 372
Db 1090 GTCCATGGGAGCGCTACGGCTCAAGTTCGATTCACGGGATCCGCCAGGCCCTCCAG 1149
QY 373 ProHisProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSer 392
Db 1150 CCCCACCCCGGAGTCACTCTGTACAGTACCCCTCAGACCTCCCGTACATGGGCTCC 1209
QY 393 TyrHisAlaHisProGlnLysMetAsnPheValAlaProHisProAlaLeuProVal 412
Db 1210 TATCACGCCACCCACAGAGATGAATTTGTGGCGCCGCCACCCCTCCAGCCCTCCCGCTG 1269
QY 413 ThrSerSerPhePheAlaAlaProAsnProTyrTyrAsnSerProThrGlyGlyIle 432
Db 1270 ACATCTCCAGTTTTTTGTGTCGCCCAACCCATAGTGGATTCCACCACTGGGGGTATA 1329
QY 433 TyrProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
Db 1330 TACCCCAACACTAGGCTCCCGACCGACCATATGCTTCTCATCTGGGCACTTACTACTAC 1386

RESULT 5
US-10-205-823-98
; Sequence 98, Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Gorbacheva, Bella
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Wonsay, Angela M.
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; THERAPY OF PROSTATE CANCER
; FILE REFERENCE: MEI-044
; CURRENT APPLICATION NUMBER: US/10/205,823
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
```

```
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,745
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 98
; LENGTH: 3166
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-205-823-98

Alignment Scores:
Pred. No.: 1,08e-238 Length: 3166
Score: 2182.00 Matches: 407
Percent Similarity: 91.72% Conservative: 14
Best Local Similarity: 88.67% Mismatches: 24
Query Match: 89.28% Indels: 14
DB: 15 Gaps: 4

US-09-902-772-2 (1-451) x US-10-205-823-98 (1-3166)
QY 2 AlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGluCys 21
Db 281 GCAGCTCATATCAGGAAGCCTTATCAGTTGTGAGTGAGGACCATGCTGTTTTCAGTGT 340
QY 22 AlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerSerGluTyr 41
Db 341 GCCTACGGAAGCCACACCTGGCTAAGACAGAGATGACCGGTCTCTCCAGCGACTAT 400
QY 42 GlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGlnProPro 61
Db 401 GGACAGATTTCCAGATGAGCCCGCTGCTCCTCAGCAGGATGGCTGTCTCAACCCCA 460
QY 62 AlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSer 81
Db 461 GCGAGGGTCACCATCAAAATGGAATGTAAACCTTAGCCAGGTGAATGGCTCAAGGAACCT 520
QY 82 ProAspAspCysSerValAlaLysGlyLysMetValSerSerSerSerAspValGly 101
Db 521 CTTGATGAATGAGTGTGGCCAAAGCGGGGAAGATGTGTGGCAGCCCGCAGACACCGTTGG 580
QY 102 MetAsnTyrGlySerTyrMetGluLysHisIleProProProAsnMetThrThrAsn 121
Db 581 ATGAACCTACGCGAGCTACATGAGGAGAGGACATGTCACCCCAAAACATCACCACCAAC 640
QY 122 GluArgArgValIleValProAlaAspProThrLeuTyrSerThrAspHisValArgGln 141
Db 641 GAGCGCAGAGTTATGTCGCCAGCATCTTACGCTATGGAGTACAGACCATGTGCGGAG 700
QY 142 TrpLeuGluTyrAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPheGln 161
Db 701 TGGCTGGAGTGGCGGTGAAGAATATAGCCCTTCCAGACGTCAACATCTTGTATTCCAG 760
QY 162 AsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThrPro 181
Db 761 AACATCGATGGGAAGAACTGTGCAAGATGACCAAGGACGACTTCCAGAGGCTCACCCCC 820
QY 182 SerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArgGlyAlaThr 201
Db 821 AGCTACAACCGGACATCTCTCTCATCATCTCCATCTCAGACGACTCTCTTCCCA 880
QY 202 PheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrThrArgPro 221
Db 881 CATTTG-----ACTTCA-----GATGATGTTGATAAAGCCTTACAAACTCTCCA 925
QY 222 -----AspLeuProTyrGluGlnAlaArgSerAlaTTP 233
Db 926 CGGTTAATGATGCTAGAAACACAGATTATGATATGAGCCCCCAGGAGATCAGCTGG 985
```

```

QY 234 ThrSerHisSerHisProThr---GlnSerLysAlaThrClnProSerSerSerThrVal 252
Db 986 ACCGGTCACGGCCACCCACCGCCCGAGTGAAGCTGCTCAACCACTCTCTTCCACAGTG 1045
QY 253 ProLysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSer 272
Db 1046 CCCAAATGAGACAGCGCTCTCAGTTAGATCCCTATCAGATTCTTGACCAACAAGT 1105
QY 273 SerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeu 292
Db 1106 AGCGGCTTGCATATCAGGAGTGGCCAGATCCAGCTTTGGCAGTTCCTCTCGAGGCTC 1165
QY 293 LeuSerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLys 312
Db 1166 CTCTCGGACAGCTCCAACTCCACCTGCATCACCCTGGAGGACCCAAACGGGAGTTCAG 1225
QY 313 MetThrAspProAspGluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMet 332
Db 1226 ATGACCGATCCCGACAGGTGGCCCGCGCTGGGGAGCGGAGAGCAACCCCAACATG 1285
QY 333 AsnTyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLys 352
Db 1286 AACTACGATAGCTACCGCGCGCTCCGTTACTACTATGACNAGAACATCATGACCAAG 1345
QY 353 ValHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGln 372
Db 1346 GTCCATGGGAAGGCTACGCTCAAGTTGCACTTCCACGGATGCCCGCCCGCTCCAG 1405
QY 373 ProHisProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSer 392
Db 1406 CCCACCCCGGAGTCACTCTGTACAGTACCCCTCAGACTCCCGTACATGGGCTCC 1465
QY 393 TyrHisAlaHisProGlnLysMetAsnPheValAlaProHisProAlaLeuProVal 412
Db 1466 TATCAGCGCCACCCACAGAGATGAACTTTGTGGCGCCCGCCCGCTCCAGCCCTCCCGTG 1525
QY 413 ThrSerSerSerPhePheAlaProAsnProTyrTrpAsnSerProThrGlyGlyIle 432
Db 1526 ACATCTTCCAGTTTTTTTGTGCGCCCAACCCCATCTGGAATTCACCACTGGGGTATA 1585
QY 433 TyrProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
Db 1586 TACCCCAACACTAGGTCCCGCCACAGCCATATGCTTCTCATCTGGGCACTTACTAC 1642

RESULT 6
US-10-021-660-26
; Sequence 26, Application US/10021660
; Publication No. US20030152926A1
; GENERAL INFORMATION:
; APPLICANT: Murray, Richard
; APPLICANT: Glyme, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: EOS Biotechnology, Inc.
; TITLE OF INVENTION: No. US20030152926A1el Methods of Diagnosis of Angiogenesis,
; TITLE OF INVENTION: Compositions and Methods of Screening for Angiogenesis
; TITLE OF INVENTION: Modulators
; FILE REFERENCE: 018501-000710US
; CURRENT APPLICATION NUMBER: US/10/021,660
; CURRENT FILING DATE: 2001-12-06
; PRIOR APPLICATION NUMBER: US/09/784,356
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: US 09/637,977
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 135
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 26
; LENGTH: 3166
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-021-660-26
Alignment Scores:
Pred. No.: 1.08e-238 Length: 3166

```

```

Score: 2182.00 Matches: 407
Percent Similarity: 91.72% Conservative: 14
Best Local Similarity: 88.67% Mismatches: 24
Query Match: 89.28% Indels: 14
DB: 15 Gaps: 4
US-09-902-772-2 (1-451) x US-10-021-660-26 (1-3166)

QY 2 AlaSerThrIleLysGluAlaLeuSerValValSerGlnSerLeuPheGluCys 21
Db 281 GCACCTCATATCAAGGAAGCTTATCAGTTGTGAGTGAGGACCAAGTCTGTTGTGAGTGT 340
QY 22 AlaTyrGlySerProHisLeuAlaLysThrGluMetThrAlaSerSerSerGluTyr 41
Db 341 GCCTACGGACCGCCACACCTGGCTAACACAGAGATGACCGCTCTCTCCAGGACTAT 400
QY 42 GlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnPro 61
Db 401 GGCAGAGACTTCCAGAGTAGGCCACCGCTCCCTCAGCAGGATTGGCTGTCTCAACCCCCA 460
QY 62 AlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSer 81
Db 461 GCCAGGCTCACCATCAAAATGGAATGTAAACCTAGCCAGGTGAATGCTCAAGGAACCTCT 520
QY 82 ProAspAspCysSerValAlaLysGlyGlyLysMetValSerSerSerSerAspAsnValGly 101
Db 521 CCTGATGAATGAGTGTGGCCAAAGGGGGAAGATGGTGGGACGCCAGACACCGCTTGGG 580
QY 102 MetAsnTyrGlySerTyrMetGluGluLysHisIleProProAsnMetThrThrAsn 121
Db 581 ATGAACTACGGCAGCTACATGGAGGAGAGACATGACCCCAACATGACCCAGCAAC 640
QY 122 GluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGln 141
Db 641 CAGCGCAGAGTTATCGTCCAGCAGATCCTACGCTATGGAGTACAGACCATGTGCGGCGAG 700
QY 142 TrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPheGln 161
Db 701 TGGCTGAGTGGCGGCGGGAAGAATATGGCTTCCAGACGCTCAACATCTGTGTATTCCAG 760
QY 162 AsnIleAspGlyLysGluLeuCysLysMetThrLysAspPheGlnArgLeuThrPro 181
Db 761 AACATCGATGGGAAGGAACGTGTCAAGATGACCAAGGACGACTTCCAGAGGCTCACGCC 820
QY 182 SerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArgGlyAlaThr 201
Db 821 AGCTACAAACCGCCAGATCCTCTCTCATCTCTCATCTCCATCTCCAGAGAGACTCTCTTCCA 880
QY 202 PheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrThrArgPro 221
Db 881 CATTTG-----ACTTCA-----GATGATGTTGATAAAGCCTTACAAAACCTCTCCA 925
QY 222 -----AspLeuProTyrGlnGlnAlaArgArgSerAlaTrp 233
Db 926 CGGTTAATGCTAGTAAACACAGATTTACCATATGAGCCCCCGCCAGAGATGACGCTGG 985
QY 234 ThrSerHisSerHisProThr---GlnSerLysAlaThrGlnProSerSerSerThrVal 252
Db 986 ACCGGTCACGGCCACCCACCGCCAGTCCGAAAGCTGCTCAACCATCTCTCTCCACAGTG 1045
QY 253 ProLysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSer 272
Db 1046 CCCAAATGAGACAGCGCTCTCAGTTAGATCCCTATCAGATTCTTGACCAACAAGT 1105
QY 273 SerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeu 292
Db 1106 AGCGGCTTGCATATCAGGAGTGGCCAGATCCAGCTTTGGCAGTTCCTCTCGAGGCTC 1165
QY 293 LeuSerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLys 312
Db 1166 CTCTCGGACAGCTCCAACTCCAGTGCATCAGCTGGGAGGACCAACCGGAGTTCAG 1225
QY 313 MetThrAspProAspGluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMet 332

```

```

Db      1226 ATGACGGATCCCGACGAGGTGGCCGCGCTGGGAGAGCGGAAGAGCAAAACCCACATG 1285
QY      |||
QY      333 AsnTyrAspLysLeuSerArgAlaLeuArgTyrTyrAspLysAsnIleMetThrLys 352
Db      1286 AACTAGATAAGCTCAGCGCGCCCTCCGTTACTACTATGACAAGAACATCATGACCAAG 1345
QY      |||
QY      353 ValHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyVileAlaGlnAlaLeuGln 372
Db      1346 GTCCATGGGAAGCGCTACGCTACAAAGTTCGACTTCCACGGGATGCCCCAGGCCCTCCAG 1405
QY      |||
QY      373 ProHisProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSer 392
Db      1406 CCCCACCCCGCGAGTCATCTGTGTAAGTACCCCTCAGACCTCCCGTACATGGGCTCC 1465
QY      |||
QY      393 TyrHisAlaHisProGlnLysMetAsnPheValAlaProHisProAlaLeuProVal 412
Db      1466 TATCAGGCCACCCACAGAGATGAATTTGTGGGCCCCACCTCCAGCCCTCCCGTG 1525
QY      |||
QY      413 ThrSerSerPhePheAlaAlaProAsnProTyrTyrAsnSerProThrGlyGlyIle 432
Db      1526 ACATCTTCCAGTTTCTTGTGCCCCAACCCCATACTGGAAATTCACAACTGGGGGTATA 1585
QY      |||
QY      433 TyrProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
Db      1586 TACCCCAACACTAGGCTCCCAACAGCCATATGCTTCTCATCTGGGCACCTACTAC 1642

```

## RESULT 7

```

US-10-087-192-1124
; Sequence 1124, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION: David W.
; APPLICANT: Morris, David W.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1124
; LENGTH: 1297
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-087-192-1124

```

```

Alignment Scores:
Pred. No.: 5,74e-225 Length: 1297
Score: 2057.00 Matches: 384
Percent Similarity: 83.69% Conservative: 11
Best Local Similarity: 81.36% Mismatches: 9
Query Match: 84.17% Indels: 68
DB: 13 Gaps: 2

```

US-09-902-772-2 (1-451) x US-10-087-192-1124 (1-1297)

```

QY      7 GluAlaLeuSerValSerGluAspGlnSerLeuPheGluCysAlaTyrGlySerPro 26
Db      2 GAGGCCITGTCAGTTGTGAGCGAGGACCAAGTCACTATTTCAGTGTGCTACGGAACGCCA 61
QY      |||
QY      27 HisLeuAlaLysThrGluMetThrAlaSerSerSerSerGluTyrGlyGlnThrSerLys 46
Db      62 CACCTGGCTAAGACAGAGATGACCGCATCTCTTCCAGTACTATGCGCCAGACATCCAAG 121
QY      |||
QY      47 MetSerProArgValProGlnGlnAspTripleuSerGlnProProAlaArgValThrIle 66
Db      122 ATGAGTCCAGAGTCCCTCAGCAGGAGTGGCTGTCTCAAGCCCAAGCCAGGGTCACCATC 181

```

```

QY      67 LysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSerProAspAspCysSer 86
Db      182 AAGATGGAGTGCAACCTAGTCAGTGAATGGTTCAGGAACCTCACCTGATGAGTGAGT 241
QY      |||
QY      87 ValAlaLysGlyLysMetValSerSerSerAspAsnValGlyMetAsnTyrGlySer 106
Db      242 GTGAACAAAGGTGGAGATGTTGGGAGCCCGGATGACTGTGGGGATGAGTACGCGAGC 301
QY      |||
QY      107 TyrMetGluGluLysHisIleProProProAsnMetThrThrAsnGluArgValIle 126
Db      302 TACATGGAGGAGAGCATGTGCGCCCTCCCAATATGACCACAAATGAGCGCAGAGTGATC 361
QY      |||
QY      127 ValProAlaAspProThrLeuTrpSerThrAspHisValArgGlnTrpLeuGluTrpAla 146
Db      362 GTCCCTTCAGATCCTACTCTGTGGAGCAGACCATGTCCGACAGTGGCTCGAGTGGCG 421
QY      |||
QY      147 VallysGluTyrGlyLeuProAspValAspIleLeuLeuPheGlnAsnIleAspGlyLys 166
Db      422 GTGAAGAAATATGGCTCTCGATGTGGAGCTTACTATTTCAGAAATATGATGGGAAG 481
QY      |||
QY      167 GluLeuCysLysMetThrLysAspAspPheGlnArgLeuThrProSerTyrAsnAlaAsp 186
Db      482 GAGCTGTGCAAGATGACAAAGGATGACTTCCAGCGGCTCAGCGGAGTACAAATGCCGAC 541
QY      |||
QY      187 IleLeuLeuSerHisLeuHisTyrLeuArgGluArg----- 198
Db      542 ATTCTTCTCTCACATCTCCACTACTCAGAGAGACTCCCTTCCACATCTGACTTCCGAT 601
QY      |||
QY      199 -----Gly 199
Db      602 GACGTTGATAAGGCTTTACAAAACCTCTCCACGGTTAATGTCATGCCAGAAACACAGGGGT 661
QY      |||
QY      200 AlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrThr 219
Db      662 GCAGCTTTATTTTCCCAATACCTTCACTATATCCGAGCTAGCAAGAAATATCAACT 721
QY      |||
QY      220 ArgProAspLeuProTyrGluGlnAlaArgSerAlaTrpThrSerHisSerHisPro 239
Db      722 AGGCCA----- 727
QY      240 ThrGlnSerLysAlaThrGlnProSerSerSerThrValProLysThrGluAspGlnArg 259
Db      727 ----- 727
QY      260 ProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGly 279
Db      728 -----GATCCTTACAGATCCTGGGACCGACAGTAGCGCTTGTCTAATCCAGGT 778
QY      280 SerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeuSerAspSerSerAsnSer 299
Db      779 AGTGCCAGATCCAGCTGTGGCAGTTCCTGTCTGAACTCCTGTGACAGAGTCCAACTCC 838
QY      300 AsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspProAspGluVal 319
Db      839 AACTGCATCACCCTGGGAAGGCCACCAACCGGGAGTTCAAGATGACAGACCCCGAGAGTG 898
QY      320 AlaArgArgTyrGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArg 339
Db      899 GCTCGGCGCTGGGGGAGAGGAAGCAAGCAACATGAATGAACTATGACAAGCTCAGCGCG 958
QY      340 AlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAla 359
Db      959 GCCCTCCCGCTACTACTACGACAAAACATCATGACCAAGGTGACCGGAGCGGCTACGCC 1018
QY      360 TyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGluSerSer 379
Db      1019 TACAAGTTTGACTTCCACGGGATTCGCCAGGCCCTTGACGCCCAACCTCTCTGAGTCGTCC 1078
QY      380 MetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisProGlnLys 399
Db      1079 CTGTACAAGTACCCCTCCGACCTGCATACATGGGCTCCTATCAGCCCAACCCCGAAG 1138
QY      400 MetAsnPheValAlaProHisProProAlaLeuProValThrSerSerSerPhePheAla 419

```

Db 1139 ATGAACCTTTGTCTCCCAACCTCCGCTCTCCAGTCACATCTTCCAGTTCTTTGCT 1198  
Qy 420 AlaProAsnProTyrTrpAsnSerProThrGlyGlyLeuValProAsnThrArgLeuPro 439  
Db 1199 TCCCGCAACCACTACTGGAATTACCGACTGGGGGCACTTCCCGAACAACACTAGGCTCCCA 1258  
Qy 440 AlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451  
Db 1259 GCCAGGCATATGCCCTCTACCTGGGCACCTACTAC 1294

RESULT 8

US-10-052-482-198  
; Sequence 198, Application US/10052482  
; Publication No. US2004007264A1  
; GENERAL INFORMATION:  
; APPLICANT: Engelhard, Eric  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER  
; FILE REFERENCE: A-71087/RMS/DCF  
; CURRENT APPLICATION NUMBER: US/10/052,482  
; PRIOR FILING DATE: 2002-08-15  
; PRIOR APPLICATION NUMBER: US 09/747,377  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 09/798,586  
; NUMBER OF SEQ ID NOS: 241  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 198  
; LENGTH: 1359  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-052-482-198

Alignment Scores:  
Pred. No.: 1.87e-175 Length: 1359  
Score: 1624.50 Matches: 306  
Percent Similarity: 79.43% Conservative: 57  
Best Local Similarity: 66.96% Mismatches: 83  
Query Match: 66.47% Indels: 11  
Gaps: 7  
US-09-902-772-2 (1-451) x US-10-052-482-198 (1-1359)

Qy 1 MetAlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20  
Db 1 ATGACGGGACTATTAAGGAGGCTCTCGTGTGTGAGCGACGACCACTCCCTTTTGC 60  
Qy 21 CysAlaTyrGly----SerProHisLeuAlaLysThrGluMetThrAlaSerSerSer 39  
Db 61 TCAGCGTACGGAGCGGCGACGCCATCTCCCAAGGCCGACATGACTGCTCGGGAGTCCT 120  
Qy 40 GluTyrGlyGlnThrSerLysMetSerProArgValProGlnAspTrpLeuSerGln 59  
Db 121 GACTACGGGAGCCCCACAGATCAACCCCTCCCAACAGCAGCAGAGTGTGATCAATCAG 180  
Qy 60 ProProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArg 79  
Db 181 ---CCAGTGGGGTCAAGTCAAGCGGAGTAT-----GACCACATGAATGATCCAGG 231  
Qy 80 AsnSerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsn 99  
Db 232 GAGTCTCCGGTGGACTGACGCGTGTAGCAATGACGAGAGTGTGGCGGAGCGAGTCC 291  
Qy 100 ValGlyMetAsnTyrGlySerTyrMetGluGlyHis----IleProProProAsnMet 118  
Db 292 AACCCCATGAACACTACACAGCTATATGACGAGAAGAATGCCCCCTCCCTCCCAACATG 351  
Qy 119 ThrThrAsnGluArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHis 138  
Db 352 ACCCAACGAGAGAGAGTATGCTGCTCCCGACAGCCCACTGTGACACAGGAGCAT 411  
Qy 139 ValArgGlnTrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeu 158

Db 412 GTGAGGCAATGGCTGGAGTGGGCCATTAAGGAGTACAGCTTGATGGAGATCGACACATCC 471  
Qy 159 LeuPheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArg 178  
Db 472 TTTTTCAGACATGGATGTCAGAGAACTGTGTAATGAACAGAGGAGGACTTCTCCCGC 531  
Qy 179 LeuThrProSerTyrAsnAlaAspIleLeuSerHisIleuHisTyrLeuArgGluArg 198  
Db 532 GCCACCACTCTTACACACGGAAGTGTGTGTCACACCTCAGTTCACCTCAGGAA-- 588  
Qy 199 GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThr 218  
Db 589 ---AGTTCACCTGCTGGCCTATATACAACTCCACACCGACCAATCTCCAGATTGAGT 645  
Qy 219 ThrArgProAspLeuProTyrGluGlnAlaArgSerAlaTrpThrSerHisSerHis 238  
Db 646 GTCAAAGAAGACCTTCTTATGACTCAGTCAGAGAGAGCTTGGGCAATAAATGATGAT 705  
Qy 239 ProThrGlnSerLysAlaThrGln---ProSerSerSerThrValProLysThrGluAsp 257  
Db 706 TCTGGCCTCAACAAAGTCTCCCTTGGAGGGGCACAAACGATCAGTAAGATACAGAG 765  
Qy 258 GlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsn 277  
Db 766 CAACGGCCCCAGCCAGATCCGTATCAGATCCTGGGGCCGACACGACGCTGCGCTAGCCAAC 825  
Qy 278 ProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeuLeuSerAspSerSer 297  
Db 826 CTTGAGAGCGGCGAGATCCAGCTGTGGCAATCTCTCTGGAGCTGCTCTCCGACAGCGCC 885  
Qy 298 AsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspProAsp 317  
Db 886 AACGCCAGCTGTATCACCCTGGAGGGGACCAACCGGGAGTTCAAAATGACGAGCCCCGAT 945  
Qy 318 GluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeu 337  
Db 946 GAGGTGGCCAGCGCTGGGGGAGCGGAAAGCAAGCCCAACATGATATTACGACAAAGCTG 1005  
Qy 338 SerArgAlaLeuArgTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArg 357  
Db 1006 AGCCGGGCGCTCCGTTATTACTATGATAAAACATATTATGACCAAGTGCAGCGCAAGA 1065  
Qy 358 TyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGlu 377  
Db 1066 TATGTTTCAAAATTTGACTTCCCGGCATTCGCCAGGCTCTGCAGCCACATCCGACCGAG 1125  
Qy 378 SerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisPro 397  
Db 1126 TCGTCCATGTACAGTACCTCTTGACATCTCTTACATGCTCTTCTACCATGCCACACAG 1185  
Qy 398 GlnLysMetAsnPheValAlaProHisProAlaLeuProValThrSerSerSerPhe 417  
Db 1186 CAGAAAGGTGAATTTGTCTCCCTCCCATCCATCTCCATGCTGCTCTCTCTCCAGTTTC 1245  
Qy 418 PheAlaAlaProAsnProTyrTrpAsnSerProThrGlyGlyIleTyrProAsn----- 435  
Db 1246 TTTGGAGCGCATCACATAGTGGACCTCCCGCGGGGGAATCTACCCCAACCCCAAC 1305  
Qy 436 ---ThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451  
Db 1306 GTCCCGCGCCATCTTAACACCCACCGTGGCTTTCACACTTAGGAGCTACTAC 1356

RESULT 9

US-10-052-482-197  
; Sequence 197, Application US/10052482  
; Publication No. US2004007264A1  
; GENERAL INFORMATION:  
; APPLICANT: Engelhard, Eric  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER  
; FILE REFERENCE: A-71087/RMS/DCF  
; CURRENT APPLICATION NUMBER: US/10/052,482



```

; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 241
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 197
; LENGTH: 2957
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-052-482-197

Alignment Scores:
Pred. No.: 5,93e-175 Length: 2957
Score: 1624.50 Matches: 306
Percent Similarity: 79.43% Conservative: 57
Best Local Similarity: 66.96% Mismatches: 83
Query Match: 66.47% Indels: 11
DB: 12 Gaps: 7

US-09-902-772-2 (1-451) x US-10-052-482-197 (1-2957)
QY 1 MetAlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeupheGlu 20
DB 173 ATGACGGGACTATTAAAGGAGCTCTCTCGGTGGTGGAGCGACGACGACCTCTTTGAC 232
QY 21 CysAlaTyrGly----SerProHisLeuAlaLysThrGluMetThrAlaSerSerSer 39
DB 233 TCACGGTACGGAGCGGCGCCACTCTCCCAAGGCGGACATGACTGCTCGGGAGTCT 292
QY 40 GluTyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTTPLeuSerGln 59
DB 293 GACTACGGGAGCGCCACAGATCAACCCCTCCACACAGCAGGAGTGGATCAATCAG 352
QY 60 ProProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArg 79
DB 353 ---CCAGTGGGGTCAAGCTCAAGCGGGAGTAT-----GACCACATGAATGATCCAGG 403
QY 80 AsnSerProAspAspCysSerValAlaLysGlyLysMetValSerSerSerAspAsn 99
DB 404 GAGTCTCCGGTGGAGTGGAGCTTAGCAATGACAGCAAGCTGGTGGCGGGAGGAGTCC 463
QY 100 ValGlyMetAsnTyrGlySerTyrMetGluLysHis----IleProProAsnMet 118
DB 464 AACCCCATGAATCAACAGCTATATGACGAGAGAAGTGGCCCTCTCTCCCAACATG 523
QY 119 ThrThrAsnGluArgArgValIleValProAlaAspProThrLeuTTPSerThrAspHis 138
DB 524 ACCACCAACGAGAGAGAGTATCGTCCCCGACAGCCCACTGTGGACACAGGAGCAT 593
QY 139 ValArgGlnTTPLeuGluTTPAlaValLysGluTyrGlyLeuProAspValAspIleLeu 158
DB 594 GTGAGGCAATGGCTGGAGTGGCCATAAAGGAGTACAGCTGTGAGGAGATCGACATCC 643
QY 159 LeupheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArg 178
DB 644 TTTTTCAGAGCATGGATGGCAGAGACTGTGTAAATGAACAAGGAGGACTTCTCCCGC 703
QY 179 LeuThrProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluArg 198
DB 704 GCCACCACCTCTACACACGGAAGTGTGTGTGACACACCTCAGTACCTACCTCAGGAA-- 760
QY 199 GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThr 218
DB 761 ---AGTTCACTGCTGGCTATATACACACCTCCACACCGGACCAATCTCCAGATTGAGT 817
QY 219 ThrArgProAspLeuProTyrGluGlnAlaArgSerAlaTTPThrSerHisSerHis 238
DB 818 GTCAAGAGAGCCCTTCTTATGACTCAGTCAAGAGAGGAGCTTGGGCAATACATGAT 877
QY 239 ProThrGlnSerLysAlaThrGln---ProSerSerSerThrValProLysThrGluAsp 257

```

```

DB 878 TCTGCCTCAACAAAAAGTCTCTCCCTTGGAGGGGCACAAACGATCAGTAAGATACAGAG 937
QY 258 GlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsn 277
DB 938 CAACGGCCCGCAGCAGATCCGATATCAGATCTCTGGGCGGACGAGCAGTCGCTAGCCAAC 997
QY 278 ProGlySerGlyGlnIleGlnLeuTTPGlnPheLeuLeuGluLeuLeuSerAspSerSer 297
DB 998 CTTGGAAGCGGGCAGATCCAGCTGTGGCAATCTCTCTGGAGTGTCTCTCCGACAGCGCC 1057
QY 298 AsnSerAsnCysIleThrTTPGluGlyThrAsnGlyGluPheLysMetThrAspProAsp 317
DB 1058 AACGCCAGCTGTATCACCTGGAGGGGACCAACCGGGAGTTTCAAAATACGAGCCCGCAT 1117
QY 318 GluValAlaArgArgTTPGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeu 337
DB 1118 GAGGTGGCCAGCGCTGGGGCGGCGAAGAACGCCCAACATGAATACGACAGCTG 1177
QY 338 SerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArg 357
DB 1178 AGCCGGGCTCTCCGTTATTACTATGATAAACATTTATGACCAAGTGCACGGCAAAAGA 1237
QY 358 TyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGlu 377
DB 1238 TATGCTTACAAATTTGACTTCACGGCATTCGACGGCTCTGCGCCACATCCGACCGAG 1297
QY 378 SerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisPro 397
DB 1298 TCGTCCATGTACAACTACCTTCTGACATCTCTACATCTCTCTACCATGCCACCAG 1357
QY 398 GlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSerSerSerPhe 417
DB 1358 CAGAAAGGTGAATTTGTCTCTCCCATCTCCATCTCCATCTCCATCTCTCTCTCTCTCT 1417
QY 418 PheAlaAlaProAsnProTyrTTPAsnSerProThrGlyGlyIleTyrProAsn----- 435
DB 1418 TTTGAGCGGCTACATACTGAGCTCCCGACGGGGGAATCTACCCCAACCCCAAC 1477
QY 436 ---ThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
DB 1478 GTCCCGCGCATCTTAACACCCAGCTGCTTCCACACTTAGGCGAGCTACTAC 1528

RESULT 10
US-10-342-887-710
; Sequence 710, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 710
; LENGTH: 2957
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-710
Alignment Scores:

```





233 TCAGCGTACGGAGCGGAGCCATCTCCCAAGCGCGACATGACTGCTCGGGAGTCT 292  
QY  
40 GluTyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGln 59  
Db  
293 GACTACCGGAGCCCAAGATCAACCCCTCCCAACAGCAGGAGTGAATCAATCAG 352  
QY  
60 ProProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArg 79  
Db  
353 ---CCAGTGAGGGTCAACGTCAGCGGAGTAT-----GACCACATGAATGATCCAGG 403  
QY  
80 AsnSerProAspAspCysSerValAlaLysGlyLysMetValSerSerAspAsn 99  
Db  
404 GAGTCTCCGGTGGACTCAGCGTTAGCAATGAGCAAGCTGTGGCGGAGCGAGTCC 463  
QY  
100 ValGlyMetAsnTyrGlySerTyrMetGluGluLysHis---IleProProAsnMet 118  
Db  
464 AACCCCATGAATACACAGCTATATGACGAGAGAGATGGCCCTCTCTCCACATG 523  
QY  
119 ThrThrAsnGluArgValIleValProAlaAspProThrLeuTyrSerThrAspHis 138  
Db  
524 ACCACCAACAGAGAGAGTATCGTCCCGCAGACCCACACTGTGGACACAGGAGCAT 583  
QY  
139 ValArgGlnTyrLeuGluTyrAlaValLysGluTyrGlyLeuProAspValAspIleLeu 158  
Db  
584 GTGAGGCAATGGCTGGAGTGGCCATTAAGAGTACAGCTGTGAGATCGACATCC 643  
QY  
159 LeuPheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArg 178  
Db  
644 TTTTTCAGAACATGATGTCGCAAGAACTGTGTAATAAGAACAGGAGACTTCTCCCG 703  
QY  
179 LeuThrProSerTyrAsnAlaAspIleLeuLeuSerHisIleuHisTyrLeuArgGluArg 198  
Db  
704 GCCACACCTCTACACAGAGAGTCTGTGTACACCTCAGTTACCTCAGCGGAA--- 760  
QY  
199 GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThr 218  
Db  
761 ---AGTTCACTGCTGGCTATATACAACTCCACACCGACCAATCTCAAGATGAGT 817  
QY  
219 ThrArgProAspLeuProTyrGluAlaArgSerAlaThrThrSerHisSerHis 238  
Db  
818 GTCAAAGAACCCCTCTTATGACTAGTCAGTACAGAGAGAGCTGTGGGCAATAACATGAAT 877  
QY  
239 ProThrGlnSerLysAlaThrGln---ProSerSerThrValProLysThrGluAsp 257  
Db  
878 TCTGGCTCAACAAAGTCTCCCTGGAGGGGACAAACGATCAGTAAGATAACAGAG 937  
QY  
258 GlnArgProGlnLeuAspProTyrGlnIleGluGlyProThrSerSerArgLeuAlaAsn 277  
Db  
938 CAACGGCCCCAGCCAGATCCGTATCAGATCTGGGCCCGACACAGCAGTCGCTAGCAAC 997  
QY  
278 ProGlySerGlyGlnIleGlnLeuTyrGlnPheLeuLeuGluLeuSerAspSerSer 297  
Db  
998 CTTGGAGCGGCGAGATCCAGCTGTGCAATCTCTCTGAGTGTCTCCGACAGGCC 1057  
QY  
298 AsnSerAsnCysIleThrTyrTyrGluGlyThrAsnGlyGluPheLysMetThrAspProAsp 317  
Db  
1058 AACGCCAGCTGTATCAGCTGGAGGGACCAACCGGGAGTTCAAAATGACGGACCCCGAT 1117  
QY  
318 GluValAlaArgArgTyrGlyGluArgCysSerLysProAsnMetAsnTyrAspLysLeu 337  
Db  
1118 GAGTGGCCAGCGCTGGGCGGCGGAAAGCAAGCCCAACATGAATACGACAGCTG 1177  
QY  
338 SerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArg 357  
Db  
1178 AGCCGGGCCCTCGTTATCTATGATATAAAACATTATGACCAAGATGACGCGCAAGA 1237  
QY  
358 TyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGlu 377  
Db  
1238 TATCTTACAAATTTGATCTCCAGCGATTCGCCAGCTTCGACGCCACATCCGACCGAG 1297  
QY  
378 SerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisPro 397  
Db  
1298 TCGTCCATGTACAGTACCCCTTCTGACATCTCTTACATGCTTCTTACCATGCCACCG 1357

398 GlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSerSerPhe 417  
Db  
1358 CAGAAGGTGAATTTGTCTCCCTCCCATCTCTCCATGCTGTGCTCTCTCCAGCTTC 1417  
QY  
418 PheAlaAlaProAsnProTyrTrpAsnSerProThrGlyGlyIleTyrProAsn----- 435  
Db  
1418 TTTGAGCGCGCATCAATACTGGACCTCCCCACGGGGGAATCTACCCCAACCCCAAC 1477  
QY  
436 ---ThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451  
Db  
1478 GTCCCCCGCATCTCTTAACACCCACGCTGCTTACACACTTAGGCGACTACTAC 1528  
RESULT 12  
US-10-007-926A-295  
; Sequence 295, Application US/10007926A  
; Publication No. US20030143539A1  
; GENERAL INFORMATION:  
; APPLICANT: BERTUCCI, FRANCOIS  
; APPLICANT: HOULGATTE, REMI  
; APPLICANT: BIRNBAUM, DANIEL  
; APPLICANT: NGUYEN, CATHERINE  
; APPLICANT: VIENS, PATRICE  
; APPLICANT: PERT, VINCENT  
; TITLE OF INVENTION: GENE EXPRESSION PROFILING OF PRIMARY BREAST CARCINOMAS  
; FILE REFERENCE: 1546-R-00  
; CURRENT APPLICATION NUMBER: US/10/007,926A  
; PRIOR FILING DATE: 2001-12-07  
; PRIOR APPLICATION NUMBER: 60/254,090  
; NUMBER OF SEQ ID NOS: 468  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 295  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: friend leukemia virus integration 1 (FLI1)  
US-10-007-926A-295  
Alignment Scores:  
Pred. No.: 5,93e-175 Length: 2957  
Score: 1624.50 Matches: 306  
Percent Similarity: 79.43% Conservative: 57  
Best Local Similarity: 66.96% Mismatches: 83  
Query Match: 66.47% Indels: 11  
DB: 15 Gaps: 7  
US-09-902-772-2 (1-451) x US-10-007-926A-295 (1-2957)  
QY 1 MetAlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerIleuPheGlu 20  
Db 173 ATGGACGGGACTATTAAAGAGGCTCTGCGTGGTGGAGCAGCAGCATCTCTTTGAC 232  
QY 21 CysAlaTyrGly---SerProHisLeuAlaLysThrGluMetThrAlaSerSerSer 39  
Db 233 TCAGCGTACGGAGCGGCGACCCATCTCCCAAGCGCGACATGACTGCTCGGGAGTCT 292  
QY 40 GluTyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGln 59  
Db 293 GACTACGGGCGAGCCCAAGATCAACCCCTCCCAACAGCAGGAGTGAATCAATCAG 352  
QY 60 ProProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArg 79  
Db 353 ---CCAGTGAGGGTCAACGTCAGCGGAGTAT-----GACCACATGAATGATCCAGG 403  
QY 80 AsnSerProAspAspCysSerValAlaLysGlyLysMetValSerSerAspAsn 99  
Db 404 GAGTCTCCGGTGGACTCAGCGTTAGCAATGAGCAAGCTGTGGCGGAGCGAGTCC 463  
QY 100 ValGlyMetAsnTyrGlySerTyrMetGluGluLysHis---IleProProAsnMet 118





```
Qy 318 GluValAlaArgAgtTtpGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeu 337
Db 1190 GAGGTGGCCAGGCGCTGGGAGAGCGGAGCAAGCCCAACATGATATGACAAAGCTG 1249
Qy 338 SerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArg 357
Db 1250 AGCGGGCCCTCCGATACTACTATGACAAACAAATATGACCAAGTGCATGCGAAAGG 1309
Qy 358 TyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGlu 377
Db 1310 TATCCCTACAGTTTGACTTCCATGGCATTCGCCAGGCCCTGCAGCCACATCCACAGAG 1369
Qy 378 SerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisPro 397
Db 1370 ACATCCATGTACAAAGTATCCCTCTGATATCTCTACATGCTTCTTACCATGCCATCAA 1429
Qy 398 GlnLysMetAsnPheValAlaProHisProAlaLeuProValThrSerSerSerPhe 417
Db 1430 CAGAGGTGACTTTGTCGGCTTCACCCATCTCTCCATGCTGTACCTCTCCAGCTTC 1489
Qy 418 PheAlaAlaProAsnProTyrTyrTyrAsnSerProThrGlyGlyIleTyrProAsn 435
Db 1490 TTTGGAGCAGCATCACAAATCTGGACCTCCGCCACTGCTGGGATCTATCCAAACCCCA 1549
Qy 436 ---ThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451
Db 1550 GTCCCCGCCATCTTAACACCCAGCTGCTTCACACTTAGGCAGCTACTAC 1600

RESULT 15
US-10-087-192-1126
; Sequence 1126, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 52945200122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1126
; LENGTH: 212231
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-087-192-1126

Alignment Scores:
Pred. No.: 9,05e-96 Length: 212231
Score: 956.50 Matches: 184
Percent Similarity: 79.75% Conservative: 9
Best Local Similarity: 76.03% Mismatches: 31
Query Match: 39.14% Indels: 18
Gaps: 13

US-09-902-772-2 (1-451) x US-10-087-192-1126 (1-212231)
Qy 227 GlnAlaArgArgSerAlaTyrThrSerHisSerHisProThrGlnSerLysAlaThr 245
Db 201508 CAGGCAAGGCGAGTTCTCTGGAGAAGTGGGTGGCCATGGACGAGACGACGATCCGCCACCCC 201567
Qy 246 -----GlnProSerSerSerThrValProLysThrGluAspG1 258
Db 201568 TGCAGCTAGGAGTGGACGGACAGCCCTGTGTGAAGGGAATGTGGCGGCACACCAATAGC 201627
Qy 258 nArgProGln-----LeuAspProTyrGlnIleLeuGlyPr 270
```

Search completed: May 28, 2004, 01:33:43  
Job time : 799 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 27, 2004, 20:54:02 ; Search time 3644 Seconds  
(without alignments)  
3695.896 Million cell updates/sec

Title: US-09-902-772-2  
Perfect score: 2444  
Sequence: 1 MASTIKKALSVSDQSLFE.....IYNTRLPAAMSHLGTYY 451

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-O=/cgn2\_1/USP01\_spool/US0902772/runat\_27052004\_162841\_2156/app\_query.fasta\_1.647  
-DB=EST -OPMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOPECL=0 -LOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR SCORE=pcpt -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTPWT=ptc -NORM=ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US0902772@cgn\_1\_3437 @runat\_27052004\_162841\_2156 -NCPU=6 -ICPU=3  
-NO MAP -LARGQUERY -NEG SCORE=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gssl:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2168.5	88.7	3064	11	AK078113	AK078113 Mus muscu
2	2157	88.3	1350	29	AY419454	AY419454 Homo sapi
3	2151	88.0	1866	11	AK050922	AK050922 Mus muscu
4	2136	87.0	1350	29	AY419456	AY419456 Mus muscu
5	1935	79.2	1350	29	AY419455	AY419455 Pan trogl
6	1613.5	66.0	1706	11	AK042111	AK042111 Mus muscu
7	1610.5	65.9	1347	29	AY418893	AY418893 Homo sapi
8	1602	65.5	3118	11	AK036855	AK036855 Mus muscu
9	1599.5	65.4	1347	29	AY418895	AY418895 Mus muscu
10	1428	58.4	1256	29	AY418894	AY418894 Pan trogl
11	1275	52.2	726	13	BU334592	BU334592 603500995
12	1123.5	46.0	709	14	CB269538	CB269538 1008445 H
13	1118.5	45.8	827	9	AJ456498	AJ456498 AJ456498
14	1055	43.2	2420	11	AK083900	AK083900 Mus muscu
15	1040	42.6	1045	12	EM456833	EM456833 AGENCOURT
16	1016.5	41.6	880	13	BQ233264	BQ233264 AGENCOURT
17	1002.5	41.0	1089	13	BQ212127	BQ212127 AGENCOURT
18	999	40.9	715	12	BG388025	BG388025 602412867
19	985	40.3	783	14	CD758623	CD758623 AGENCOURT
20	984	40.3	949	12	BI558012	BI558012 603240720
21	984	40.3	642	12	BI713036	BI713036 ie01d08.Y
22	984	40.3	668	10	B6660034	B6660034 B6660034
23	980	40.1	592	12	BM088063	BM088063 501436 MA
24	975.5	39.9	915	12	BG390291	BG390291 602416255
25	971.5	39.8	701	9	AU136709	AU136709 AU136709
26	970	39.7	629	10	AW948986	AW948986 QV4-FT000
27	966	39.5	1074	13	BQ953977	BQ953977 AGENCOURT
28	961	39.3	606	13	BU478556	BU478556 603471895
29	948	38.8	675	10	B633043	B633043 B633043
30	947.5	38.8	634	12	BM489636	BM489636 pgmch.BK0
31	941	38.5	616	14	CB579500	CB579500 AMGNNUC.N
32	938.5	38.4	976	13	BX328036	BX328036 BX328036
33	935.5	38.3	928	12	BG259459	BG259459 602378556
34	932	38.1	602	14	CB582161	CB582161 AMGNNUC.N
35	932	38.1	629	10	B5577887	B5577887 B5577887
36	886	36.3	633	13	EX100094	EX100094 EX100094
37	872.5	35.7	526	14	CA405791	CA405791 1001940 H
38	853.5	34.9	521	13	BU440626	BU440626 604144050
39	846	34.6	802	14	CB952600	CB952600 AGENCOURT
40	840.5	34.4	671	12	BI558530	BI558530 603240526
41	830	34.0	1201	9	AL554108	AL554108 AL554108
42	828.5	33.9	666	12	BG390572	BG390572 602416344
43	822	33.6	553	12	BG258664	BG258664 602471114
44	821	33.6	851	13	BU108155	BU108155 603110416
45	816.5	33.4	524	9	AI794504	AI794504 fc45b11.Y

# ALIGNMENTS

RESULT 1

AK078113 3064 bp mRNA linear HTC 18-SEP-2003  
LOCUS Mus musculus adult male medulla oblongata cDNA, RIKEN full-length  
DEFINITION enriched library, clone:630501C04 product:ERG mRNA, mouse homolog  
of Human ets-related gene ERG, transcript variant 1, full insert  
sequence.

ACCSSSTON AK078113  
VERSION AK078113.1 GI:26346964  
KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



Qy	181	ProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGluA-g-----	198
Db	601	CCGAGCTACAAATGCCGACATTCTTTCTCTCACATCTCCACTACCTCAGAGAGACTCCGCCCT	660
Qy	198	-----	198
Db	661	CCACATCTGACTCCGATGACGTTGATAGGCTTTACAAAACCTCTCACGGTTAATGCAT	720
Qy	199	-----GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAa	213
Db	721	GCCAGAAACACAGGGGGTGACGCTTTATTATTTCCCAAATATCTTCAGTATATATCCGAAGCT	780
Qy	214	ThrGlnArgIleThrThrArgProAspLeuProTyrGluGlnAlaArgArgSerAlaTyr	233
Db	781	ACGCAAGAAATTACACTAGGCCA-----	804
Qy	234	ThrSerHisSerHisProThrGlnSerLysAlaThrGlnProSerSerSerThrValPro	253
Db	805	-----GCTGCTCAGCCATCTCCCTCTCGAGTCGCC	834
Qy	254	LysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSer	273
Db	835	AAAACTGAAGACCGAGCTCTCTCAGTTAGATCTTTACCAGATCTCTGGACCGACGATGAC	894
Qy	274	ArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTyrGlnPheLeuLeuGluLeuLeu	293
Db	895	CGCTTCTGCTAATCCAGGTAGTGGCCAGATCCAGCTGTGGCAGTCTCTGCTCGAACCTCTG	954
Qy	294	SerAspSerSerAsnSerAsnCyfIleThrTyrGluGlyThrAsnGlyGluPheLysMet	313
Db	955	TCACACAGCTCAAACTCCAACTGCATGCATCACCTGGGAGAGCCAAACGGGGAGTTCAAGATG	1014
Qy	314	ThrAspProAspGluValAlaArgTyrGluArgLysSerLysProAsnMetAsn	333
Db	1015	ACAGACCGGACGAGGTGGCTCGCGCTGGGGGAGAGAGAGCAAGACCAACATCATGAC	1074
Qy	334	TyrAspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysVal	353
Db	1075	TATGACAAGCTCAGCGCGCCCTCCGCTACTACTACGACAAAAACATCATGCCAAGGTG	1134
Qy	354	HisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnPro	373
Db	1135	CACGGGAGCGCTACGCTCAAGATTTGACTTCCACGGGATGTCGCCAGGCCCTGCAGCCC	1194
Qy	374	HisProProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyr	393
Db	1195	CACCTCTCTGAGTCGTCCTGTACAAGTACCCCTCCGACCTGCCATACATATGGGCTCTAT	1254
Qy	394	HisAlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProValThr	413
Db	1255	CACCCACACCCCCAGAGATGAATTTGTCTCCCAACCCCTCCCGCTCTCCAGTCAACA	1314
Qy	414	SerSerSerPhePheAlaAlaProAsnProTyrTyrTrpAsnSerProThrGlyGlyIleTyr	433
Db	1315	TCCTTCAGATTCTTTGCTTCTCCCGACCCATACCTAGGAATTCACCGACTGGGGCATCTAC	1374
Qy	434	ProAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr	451
Db	1375	CCGAACACTAGGCTCCCGAGCCGCAATATGCCCTCTCACCTGGGCACCTACTAC	1428
RESULT 2			
LOCUS	AY419454	1350 bp	DNA linear
DEFINITION	Homo sapiens ERG gene, VIRTUAL TRANSCRIPT, partial sequence,		
ACCESSION	AY419454		
VERSION	AY419454.1		
KEYWORDS	GSS		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	1 (bases 1 to 1350)		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
	GSS 17-DEC-2003		

**AUTHORS**  
 Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,  
 Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,  
 Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,  
 Adams,M.D. and Cargill,M.

**TITLE**  
 Inferring nonneutral evolution from human-chimp-mouse orthologous  
 gene trios

**JOURNAL**  
 Science 302 (5652), 1960-1963 (2003)

**PUBMED**  
 14671302

**REFERENCE**  
 2 (bases 1 to 1350)

**AUTHORS**  
 Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,  
 Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,  
 Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,  
 Adams,M.D. and Cargill,M.

**TITLE**  
 Direct Submission

**JOURNAL**  
 Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
 Rockville, MD 20850, USA

**COMMENT**  
 These sequences were made by sequencing genomic exons and ordering  
 them based on alignment.

**FEATURES**  
 Location/Qualifiers  
 source  
 1..1350  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 gene  
 <1..>1350  
 /gene="ERG"  
 /locus\_tag="HCM6896"

**ORIGIN**  
 Alignment Scores:  
 Pred. No.: 4,616-176 Length: 1350  
 Score: 2157.00 Matches: 400  
 Percent Similarity: 91.09% Conservative: 9  
 Best Local Similarity: 89.09% Mismatches: 36  
 Query Match: 88.26% Indels: 4  
 DE: 29 Gaps: 2

US-09-902-772-2 (1-451) x AY419454 (1-1350)

Qy 7 GluAlaLeuSerValValSerGluAspGlnSerLeuPheGluCysAlaTyrGlySerPro 26  
 Db 1 GRAGCCCTTATCAGTTGTGAGTGAGGACCAGTCGTTGTTGAGTGTGCTACCGAACGCCA 60  
 Qy 27 HisLeuAlaLysThrGluMetThrAlaSerSerSerSerGluTyrGlyGlnThrSerLys 45  
 Db 61 CACCTGGCTTAAGACAGAGATACCGCGTCCTCCAGCGCATATGACAGACATTCGAAG 120  
 Qy 47 MetSerProArgValProGlnAlaAspTrpLeuSerGlnProProAlaArgValThrIle 66  
 Db 121 ATGAGCCCAACGGTCCCTCAGCAGGATGGCTGTCTCAACCCACGACGAGGTCCACATC 180  
 Qy 67 LysMetGluCysAsnProAsnGlnValasnGlySerArgAsnSerProAspAspCysSer 86  
 Db 181 AAAATGGAAATGTAACCTTAGCCAGTGTAATGGCTCAAGGAACCTCTCTGTATGAATGCAGT 240  
 Qy 87 ValAlaLysGlyGlyLysMetValSerSerSerAspAsnValGlyMetAsnTyrGlySer 106  
 Db 241 GTGGCCNAAGGGCGGAAGANTGGTGGGACGCCAGACACCGTTGGATGAATACGGCAGC 300  
 Qy 107 TyrMetGluGluLysHisIleProProProAsnMetThrThrAsnGluArgArgValIle 126  
 Db 301 TACATGGAGGAGAAGACATATGCACCCCAACATGACCAACGACGACGCGCAGAGTTATC 360  
 Qy 127 ValProAlaAspProThrLeuTrpSerThrAspHisValArgGlnTrpLeuGluTrpAla 146  
 Db 361 GTGCCAGCAGATCTTACGCTATGGNATACAGACCATGTGCGGCAAGTGGCTGGAGTGGCG 420  
 Qy 147 ValLysGluTyrGlyLeuProAspValAspIleLeuLeuPheGlnAsnIleAspGlyLys 166  
 Db 421 GTGAAGAATAATGGCTTCCAGACGCTCAACATCTTTATTCACAGAACATCTCGATGGGAG 480  
 Qy 167 GluLeuCysGlySerMetThrLysAspAspPheGlnArgLeuThrProSerTyrAsnAlaAsp 186  
 Db 481 GRACTGTGAAGATGACCAAGACACATTCAGAGGCTCACCCCAAGCTTAACCGCGAC 540





CDs

```

/clone="D030036124"
/clone_type="whole body"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="9 days embryo"
249..1637
/notes="unnamed protein product; Erg mRNA, mouse homolog of
Human ets-related gene ERG, transcript variant 1
(DBJ|AB073078, evidence: BLASTN, 99%, match=1768)
putative"
/codon_start=1
/protein_id="BAC34461.1"
/db_xref="GI:26341598"
/translation="MIQVDPDAHIKELSVSEDSQLEPCAVGTEPHLAKTEMTASS
SSDYGTSKMSPRVPOCDMLSOAPARVITKACNPSQVGS RNSPDECSYKSGKMG
SPDVGMGYSYMEKHYVPBNMTNERVAVPADPTLMSIDHVRQMLEAVKYGIL
DVIDLIPNDIKXELCKMTKDFORLTPSVNADILSLHLYRETPLPLTSDVDKA
LQNSPRLMHANTDLPYPPRSAMTGHSHLTPOSKAQSPSPAVPTTEPDQOLDPY
QILGETSRILNPSGGIQLMQFLBELISBSNCTTWSTNGEPFMTTPDVARV
GERSKPMWNYDKLSRALRYIDKNITKVGKRYAKTFDHGIAQLQHPRESSLY
KYPEDLPYMGSYHAHPQKMFVSPHPALPVTSFPASBPYWNSTPTGIIYNTRLP
ASHMPSHLGIYY"

```

## ORIGIN

Alignment Scores:

Pred. No.:	2,51e-175	Length:	1866
Score:	2151.00	Matches:	402
Percent Similarity:	91.07%	Conservative:	16
Best Local Similarity:	87.58%	Mismatches:	27
Query Match:	88.01%	Indels:	14
DB:	11	Gaps:	4

US-09-902-772-2 (1-451) x AK050922 (1-1866)

```

QY 2 A1aserThrlleYgAlaleuSeYValaSerGluaspGlnSerLeuPheGluCys 21
DB 273 GCAGCTCATATTAAAGAGCGCTTGCAGTTGAGACGAGACCATCTATTGAGAGT 332
QY 22 A1aTYrGlySerProHleuAlaYerThrGluMetThrAlaSerSerSerGluTYr 41
DB 333 GCCTACGAGACGCGCACCTGGCTAAGACAGAGATGACCGCATCTCTCCAGTGCAT 392
QY 42 GlyGlnThrSerLeuMetSerProArgValProGlnGlnAspTrpLeuSerGlnProPro 61
DB 393 GCGCAGACATCCAAATAGATGATCCAGATCCTCCAGAGAGATGAGTCTGTCACAGCCCA 452
QY 62 AlaArgValThrIleYMetGluCysAsnProAsnGlnValAsnGlySerThArgAsnSer 81
DB 453 GCCAGAGGTACACCATAGATGAGATGCAACCTACTGAGTGAAGTGAAGTTCACAGAACTCA 512
QY 82 ProAspAspCysSerValAlaYSerGlyLeuMetValSerSerSerAspAsnValGly 101
DB 513 CCTGATGAGTGAAGTGAAGCAAAAGTGGAGATGCGGGAGCCCGGATATGATGGGG 572
QY 102 MetAsnTYrGlySerTYrMetGlnGluYshIleProProProAsnMetThrThrasn 121
DB 573 ATGAGCTACGCGACGCTACATGAGAGAGACATGCGCCGCCCAATATGACCAAAAT 632
QY 122 GluArgArgValIleValProAlaAspProThrLeuTrpSerThAspHisValArgGln 141
DB 633 GAGCGGAGGTATGCTCCCTGCAATCTTCTTCTGAGACAGACCAATGTCGAGAG 692
QY 142 TrpLeuGluTrpAlaValYSerGlyTYrGlyLeuProAspValAspIleleuLeuPheGln 161
DB 693 TGGCTGAGTGGCGGTGAAGAAATATGCGCTCTCGATGATGAGAGCTTACTATTAG 752
QY 162 AsnIleAspGlyYleGluLeuCYlyMetThrIlyAspAspPheGlnArgLeuThrPro 181
DB 753 AATATGAGAGGAGAGAGCTGTGACAGATGACAAAGAGATCTCCAGGCGCTCACGGCG 812
QY 182 SerTYrAsnAlaAspIleleuLeuSerHisleuHisTYrLeuArgGlnArgGlyAlaThr 201
DB 813 AGCTTAATGCGGACATTTCTTCTCTACATCTCCACTACAGAGAGACTCCCGCTCCA 872

```

```

QY 202 PheIlePheProAsnThrSerValTYrProGlnAlaThrGlnArgIleThrThArgPro 221
DB 873 CATCTG-----ACTTCC-----GATGACGTTGATTAAGGCTTACAAACTCTCCA 917
QY 222 -----AspLeuProTYrGluGlnAlaArgArgSerIleATP 233
DB 918 CGGTTAATGCATGCCAGAAACACAGATTACTTATGAGCTTCCAGAGATCAGCTGAG 977
QY 234 ThrSerHisSerHisProThr---GlnSerIysAlaThrGlnProSerSerSerThrVal 252
DB 978 ACCGGCCACAGCCACCTCATCCCTCAGTCCAAAGCTGCTCAGCCATCTCCCTCTGACAGT 1037
QY 253 ProLYeThrGluAspGlnArgProGlnLeuAspProTYrGlnIleleuGlyProThrSer 272
DB 1038 CCCAAACTGAAGACACAGGCTCCTCAGTTAGATCCTTACACATCTGAGACCAGCAAT 1097
QY 273 SerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGlnLeu 292
DB 1098 AGCCGCTTGTCTATTCAGATGAGTGGCCAGATCCAGCTGTGGAGATTCCTGCTCGAATC 1157
QY 293 LeuSerAspSerSerAsnSerAsnCYleIleTrpGluGlyThrasnGlyGluPheIys 312
DB 1158 CTGTCAAGACAGCTCCCACTCCAACTGCATCAGCTGGAGAGCACCAACGAGGAGTTCAAG 1217
QY 313 MetThraspProAspGluValAlaArgArgTrpGlyGluArgIysSerIysProAsnMet 332
DB 1218 ATACACAGCCCGACAGAGTGGCTCGGCTGGGGGAGAGAGAGACCAAGCCCAATG 1277
QY 333 AsnTYrAspIysLeuSerArgAlaLeuArgTYrTYrTYrAspIysAsnIleMetThrIys 352
DB 1278 AACTATGACAGAGCTCAGCCGCGCTCCGCTACTACTACAGCAAAACATCATGACCAAG 1337
QY 353 ValHisGlyYleAsnArgTYrAlaTYrIlysPheAspPheHisGlyIleAlaGlnAlaLeuGln 372
DB 1338 GTGCACAGGAGAGGCTACAGCTTCAAGATTTGACTTCACAGGATTTCCACAGCCCTCGAG 1397
QY 373 ProHisProProGluSerSerMetTYrIysTYrProSerAspLeuProTYrMetSerSer 392
DB 1398 CCCCACTCTCTAGTCTGCTTCCGTAAGATACCTCCGACCTGCAATACAGGGCTCC 1457
QY 393 TyrHisAlaHisProGlnIysMetAsnPheValAlaProHisProProAlaLeuProVal 412
DB 1458 TATCACGCCACCCACCAAGAGATGAACTTGATCTCCCACTCCGCTCCCAATG 1517
QY 413 ThrSerSerSerPhePheAlaAlaProAsnProTYrTrpAsnSerProThrGlyGlyIle 432
DB 1518 ACACTTCCAGTTCTTCTTCTTCCCGAACCATCTGGAATTCACAGACTGGGGGATC 1577
QY 433 TyrProAsnThrArgLeuProAlaAlaHisMetProSerHisleuGlyThrTYrTYr 451
DB 1578 TACCCGAACTAGAGCTCCAGCCAGACCATATGCTCCCTCACTGAGCACTACTAC 1634

```

RESULT 4  
AY419456 1350 bp DNA linear GSS 17-DEC-2003  
LOCUS AY419456  
DEFINITION Mus musculus ERG gene, VIRUAL TRANSCRIPT, partial sequence,  
genomic survey sequence.  
ACCESSION AY419456  
VERSION AY419456.1 GI:39775413  
KEYWORDS GSS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Taxonomy: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarirai,A.,  
Todd,M.A., Tanenbaum,D.M., Civeille,D.R., Lu,F., Murphy,B.,  
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sinsky,J.J.,  
Adams,M.D. and Carball,M.  
TITLE  
JOURNAL Science 302 (5652), 1960-1963 (2003)  
PubMed 14671302  
Interfering nonneutral evolution from human-chimp-mouse orthologous  
gene tritos

REFERENCE 2 (bases 1 to 1350)  
 AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,  
 Todd,M.A., Tanenbaum,D.M., Civeille,D.R., Lu,F., Murphy,B.,  
 Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Smirsky,J.J.,  
 Adams,M.D. and Cargill,M.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
 Rockville, MD 20850, USA  
 COMMENT These sequences were made by sequencing genomic exons and ordering  
 them based on alignment.  
 FEATURES  
 source  
 1..1350  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 <1..>1350  
 /gene="ERG"  
 /locus\_tag="HGM6896"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 2,236-173 Length: 1350  
 Score: 2126.00 Matches: 395  
 Percent Similarity: 90.42% Conservative: 11  
 Best Local Similarity: 87.97% Mismatches: 39  
 Query Match: 86.99% Indels: 4  
 Gaps: 2  
 DB:  
 US-09-902-772-2 (1-451) x AY19456 (1-1350)  
 QY 7 GUAUAAUSeValaValaSerGluuApGlnSerLeuPheGluCyAlaTyrGlySerPro 26  
 Db 1 GAGGCTTGTCAGTGTGAGCGAGGACCAAGTCACATTTGAGTGTGCTTACGAAACCCA 60  
 QY 27 HisLeuAlaIyThGluMetThzAlaSerSerSerSerGlyTyrGlyGlnThSerLys 46  
 Db 61 CACCTGGCTAAAGACGAGATGACCCGATCTTCCAGTGAATAGGCGAACAATCCAAAG 120  
 QY 47 MetSerProArzValaProGlnGlnAspTyrLeuSerGlnProAlaArgValThLe 66  
 Db 121 ATGAGTCCCAAGAGTCCCTCAGAGAGAGCTGGCTGTCTCAAGCCCAAGGAGTCAACATC 180  
 QY 67 LysMetGluCyAsnProAsnGlnValAsnGlySerArgAsnSerProAspAspCysSer 86  
 Db 181 AAGATGGAGTGCACACCTAGTCAAGTGAATGTTCCAGAACTCACTGATGAGTGAAGT 240  
 QY 87 ValAlaIySgIyGlyLysMetValaSerSerSerSerSerSerValaGlyMetAsnTyrGlySer 106  
 Db 241 GTGAACAAAGGTGGGAAGATGTGGGACCCCGATCTGTGGGATGAGCTACGCGACG 300  
 QY 107 TyrMetGluGluLysHisIleProProProAsnMetThrThraAsnGluArgArgValIle 126  
 Db 301 TACATGGAGGAGAAACATGTGCCCGCTCCCAATATAGCAACAATGAGCGCAGAGTATC 360  
 QY 127 ValProAlaAspProThleuThrSerThraSphIleValaGlyGlnTyrLeuGluTyrAla 146  
 Db 361 GTCCCTGGAGATCTTCTGTGGAGCAACAGCTGTCACAGTGGCTGAGTGGGCG 420  
 QY 147 ValLysGluTyrGlyLysProAspValaIleLeuLeuPheGlnAsnIleAspGlyLys 166  
 Db 421 GTGAAGAATATGGCTCCTCGATGTGGACGCTTACTATTTCAGAAATATGATGAGGAG 480  
 QY 167 GluLeuCyLysMetThrLysAspAspPheGlnGlyLeuTyrProSerTyrAsnAlaAsp 186  
 Db 481 GAGCGTGTCAAGATACAAAGATACCTTCAGCGGCTCAACGCGAGTCAAAATGCCGAC 540  
 QY 187 IleLeuLeuSerHisIleuHisIleuThraArgLys-----ArgGlyAlaThrPheIle 203  
 Db 541 ATTCTTCTCTCACTCTCCACTCACTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600  
 QY 204 PheProAsnThSerValTyrProGluAlaThrGlnArgIleThThArgProAspLeu 223  
 Db 601 KNNNA 660

QY 224 ProTyrGluGlnAlaArgSerAlaTyrThrSerHisSerHisProThr--GlnSer 242  
 Db 661 CCTTATGAGCCTCCCGAGGAGATGAGCTGGACCGGCGACAGCACCCTCCCTCAGTCC 720  
 QY 243 LysAlaIleThGlnProSerSerSerThraValProLysThGluAspGlnArgProGlnLeu 262  
 Db 721 AAGCTGCTCAGCATCTCTCTGTGAGTGGCCCAAACTGAAAGACCAAGGCTCTCAGTTA 780  
 QY 263 AspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGlySerGlyGln 282  
 Db 781 GATCCTTACCAATCTCTGGGACCGGACAGTACCCCTGTGTAATCCAGATGTGGCCAG 840  
 QY 283 IleGlnLeuTyrGlnPheLeuLeuGluLeuLeuSerAspSerSerAsnSerAsnGlyIle 302  
 Db 841 ATCCAGCTGTGGAGTCTCTGCTCAACTCTGTCAAGACGCTCAACTCCAACTCCATC 900  
 QY 303 ThrTyrGluGlyThraAsnGlyLysPheLysMetThrAspProAspGluValAlaArgArg 322  
 Db 901 ACCTGGAGAGGACCAACGAGGAGTTCAAGATGACAGACCCGACGAGGTGCTCGGCGC 960  
 QY 323 TyrGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArgAlaLeuArg 342  
 Db 961 TGGGGGAG 1020  
 QY 343 TyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaTyrLysPhe 362  
 Db 1021 TACTACTACGACAAACATCATGACCAAGGTGCAAGGAGAGAGAGAGAGAGAGAGAGTT 1080  
 QY 363 AspPheHisGlyIleAlaGlnAlaLeuGlnProHisProProGluSerSerMetTyrLys 382  
 Db 1081 GACTTCACAGGAGATGCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
 QY 383 TyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisProGlnLysMetAsnPhe 402  
 Db 1141 TACCTCTCCGACCTGTCATACATGGCTCTTATCAGCCCAACCCCGAAGATGAAGTTT 1200  
 QY 403 ValAlaProHisProProAlaLeuProValThrSerSerSerPhePheAlaAlaProAsn 422  
 Db 1201 GTGTCTCCCAACCTCCGCTCCCAAGTCAATCTTCCAGTTCTTTCTTCCCGAGAC 1260  
 QY 423 ProTyrTyrAsnSerProThrGlyLysIleTyrProAsnThraArgLeuProAlaIleHis 442  
 Db 1261 CCATCTGGAATTCACCGAGCTGGGGCATCTACCCGAACTAGAGCTCCACGACGACCAT 1320  
 QY 443 MetProSerHisLeuGlyThrTyrTyr 451  
 Db 1321 ATGCCTCTCACTGGGACCACTATCAG 1347  
 Db  
 RESULT 5  
 AY19455  
 LOCUS  
 DEFINITION Pan troglodytes ERG gene, VIRTUAL TRANSCRIPT, partial sequence,  
 genomic survey sequence.  
 ACCESSION AY19455  
 VERSION AY19455.1 GI:39775412  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Pan troglodytes (chimpanzee)  
 Pan troglodytes  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.  
 REFERENCE  
 1 (bases 1 to 1350)  
 Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,  
 Todd,M.A., Tanenbaum,D.M., Civeille,D.R., Lu,F., Murphy,B.,  
 Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Smirsky,J.J.,  
 Adams,M.D. and Cargill,M.  
 TITLE  
 JOURNAL Science 302 (5652), 1960-1963 (2003)  
 PUBMED 14671302  
 REFERENCE  
 2 (bases 1 to 1350)  
 Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,  
 Todd,M.A., Tanenbaum,D.M., Civeille,D.R., Lu,F., Murphy,B.,

Perriera, S., Wang, G., Zheng, X.H., White, T.J., Smirsky, J.J., Adams, M.D. and Cargill, M.

Direct Submission

Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA

COMMENT These sequences were made by sequencing genomic exons and ordering them based on alignment.

FEATURES Location/Qualifiers

1..1350

/organism="Pan troglodytes"

/mol\_type="genomic DNA"

/db\_xref="taxon:9598"

<1..>1350

/gene="ERG"

/locus\_tag="HCM6896"

ORIGIN

Alignment Scores:

Pred. No.: 7,86e-157 Length: 1350  
Score: 1935.00 Matches: 367  
Percent Similarity: 83.74% Conservative: 9  
Best Local Similarity: 81.74% Mismatches: 69  
Query Match: 79.17% Indels: 4  
D: 29 Gaps: 2

US-09-902-772-2 (1-451) x AY19455 (1-1350)

QY 7 GUAAlaLeuSerValValSerGluAapGlnSerLeuPheGluCysAlaTyrGlySerPro 26  
DB 1 GAAGCCTTATCAGATTGTGAGCGAGGACCACTGTTGATGTGCTCAGCGAAACGCGCA 60  
QY 27 HsLeuAlaLeuThrGluMetThrAlaSerSerSerSerGlyTyrGlyGlnThrSerLys 46  
DB 61 CACCTGGCTAAACAGAAATGACCGCGTCTCTCCAGGACTATGAGACAGCTTCCAG 120  
QY 47 MetSerProAryValProGlnGlnAapTyrLeuSerGlnProProAlaAryValThrIle 66  
DB 121 ATGAGGCCACGCGCTCCCTCAGAGGATGCTGCTCAACCCCGCAGCGCTCACCTC 180  
QY 67 LysMetGluCysAapProAangInValAangLysAryAangSerProAapApcCysSer 86  
DB 181 AAAATGGAATGTAACTTACCGAGGTGACCGCTCAAGNNNNNNNNNNNNNNNNNNNNNN 240  
QY 87 ValAlaLysGlyGlyLysMetValSerSerSerAapAenValGlyMetAenTyrGlySer 106  
DB 241 GTGGCCAAAGCGCGGAGATGTGGGCAAGCCAGACACCGTTGGATGAACTACGCGAGC 300  
QY 107 TyrMetGluGluLysHisIleProProProAenMetThrThrAangLysAryValIle 126  
DB 301 TACATGAGAGAGAGACATGCGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 360  
QY 127 ValProAlaAapProThrLeuTyrSerThrAapHisValAryGlnTyrLeuGlnTyrAla 146  
DB 361 NNN 420  
QY 147 ValLysGluTyrGlyLeuProAapValAapIleLeuLeuPheGlnAapIleAapLys 166  
DB 421 GGAAAGAAATATGGCTTCCAGNNNTCAATCATNNNTTATTCAGAAACATCATGGGAAN 480  
QY 167 GluLeuCysLysMetThrLysAapAapPheGlnAryGluLeuThrProSerTyrAapAlaAap 186  
DB 541 ATCTCTCTCATCATCTCATCTCACTCACTCAAGAGANNNNNNNNNNNNNNNNNNNNNN 600  
QY 187 IleLeuLeuSerHisLeuHisTyrLeuAryGlu-----AryGlyAlaThrPheIle 203  
DB 541 ATCTCTCTCATCATCTCATCTCACTCACTCAAGAGANNNNNNNNNNNNNNNNNNNNNN 600  
QY 204 PheProAenThrSerValTyrProGluAlaThrGlnAryGleThrThrAryProAapLeu 223  
DB 601 NNN 660  
QY 224 ProTyrGluGlnAlaAryArySerAlaTyrThrSerHisSerHisProThr--GlnSer 242

DB 661 CCATATGAGCCCCCGCAGAGATCAGCTGAAACCGGTCAGAGGCCACCCAGCCCGAGTCG 720  
QY 243 LysAlaThrGlnProSerSerSerThrValProLysThrGluAapGlnAryProGlnLeu 262  
DB 721 AAAGCTGCTCAACCATCTCTCCACAGGCGCCCAAAACGAAACAGAGGCTCTCAGTTA 780  
QY 263 AspProTyrGlnIleLeuGlyProThrSerSerAryGluAlaAapProLysSerGlyGln 282  
DB 781 GATCCTTATCAGATTTCTTGGACCAACAGATAGCCGCTTGCAATACAGGAGTGGCCAG 840  
QY 283 IleGlnLeuTyrGlnPheLeuLeuGluLeuLeuSerAapSerSerAapSerAenCysIle 302  
DB 841 ATCCAGCTTGGAGATTCCTCCCTGAGAGCTCTCTGCGAAGAGCTCCCACTCCACTCATC 900  
QY 303 ThrTyrGluGlyThrAangLysIlePheLysMetThrAapProAapGluValAlaAryAry 322  
DB 901 ACCGGAGAGGACCAACGGGGAGTTCAAGATACGATCCGACAGAGTGGCCCGGCGC 960  
QY 323 TrrGlyLysLysSerLysProAenMetAenTyrAapLysLeuSerAryAlaLeuAry 342  
DB 961 TGGGAGAGCGGAGAGGAAACCAACATGAACTACGATAGCTCAGCGCGCTCCGCT 1020  
QY 343 TyrTyrTyrAapLysAenIleMetThrLysValHisGlyLysAryGlyTyrAlaTyrLysPhe 362  
DB 1021 TACTACTATGACAAAGACATCATGACCAAGTCCATGGAGAGCGCTACGCTACAGATTC 1080  
QY 363 AspPheHisGlyIleAlaGlnAlaLeuGlnProHisProProGluSerMetTyrLys 382  
DB 1081 GACTTCCACGGGATTCGCCAGGCGCTTCAGCCCAACCCCGGAGATCTCTGTACAAG 1140  
QY 383 TyrProSerAapLeuProTyrMetSerSerTyrHisAlaHisProGlnLysMetAapPhe 402  
DB 1141 TACCCCTCAAGACCTCCGTCATGAGGCTCTTACAGGCCCAACCAAGAGAACTT 1200  
QY 403 ValAlaProHisProProAlaLeuProValThrSerSerSerPhePheAlaAapAapAen 422  
DB 1201 GTGGCGCCCAACCTCCAGGCGCTCCCGTACATCTTCCAGTTTTCGTGCCCAAC 1260  
QY 423 ProTyrTyrAapSerProThrGlyGlyIleTyrProAenThrAryLeuProAlaAapHis 442  
DB 1261 CCATACGTAATTCACCAACTGGGGATATACCCCAATCTAGGCTCCCGACACGACAT 1320  
QY 443 MetProSerHisLeuGlyThrTyrTyr 451  
DB 1321 ATGCCCTTCATCTGGGCACTTACTAC 1347

RESULT 6  
AK042111  
LOCUS  
DEFINITION Mus musculus 3 days neonate thymus cDNA, Riken full-length enriched library, clone: A63005911 product: Friend leukemia integration 1, full insert sequence.  
ACCESSION AK042111  
VERSION AK042111.1 GI:26334936  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
1 Carninci, P. and Hayashizaki, Y.  
TITLE High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE 9879253  
PubMed 10349636  
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
JOURNAL 20499374  
MEDLINE 11042159  
PubMed

```

3  Shiba.K., Itoh.M., Aizawa.K., Nagaoka.S., Sasaki.N., Carninci.P.,
   Konno.H., Akiyama.J., Nishi.K., Kitsuai.T., Tashiro.H., Itoh.M.,
   Sumi.N., Ishii.Y., Nakamura.S., Hazama.M., Nishine.T., Harada.A.,
   Yamamoto.R., Matsumoto.H., Sakaguchi.S., Ikegami.T., Kashiwagi.K.,
   Fujiwaka.S., Inoue.K., Togawa.Y., Izawa.M., Ohara.E., Watahiki.M.,
   Onedaki.Y., Ishikawa.T., Ozawa.K., Tanaka.T., Matsuda.S., Kawai.J.,
   Kozaki.Y., Muramatsu.M., Inoue.Y., Kira.A. and Hayashizaki.Y.
   RIKEN integrated sequence analysis (RISA) system-384-format
   sequencing pipeline with 384 multicapillary sequencer
   Genome Res. 10 (11), 1757-1771 (2000)
20530913
JOURNAL MEDLINE
PUBMED
11076861
4  The RIKEN Genome Exploration Research Group Phase II Team and the
   FANTOM Consortium.
   Functional annotation of a full-length mouse cDNA collection
   Nature 409, 685-690 (2001)
5  The FANTOM Consortium and the RIKEN Genome Exploration Research
   Group Phase I & II Team.
   Analysis of the mouse transcriptome based on functional annotation
   of 60,770 full-length cDNAs
   Nature 420, 563-573 (2002)
6  (bases 1 to 1706)
   Adachi.J., Aizawa.K., Akimura.T., Arakawa.T., Bono.H., Carninci.P.,
   Fukuda.S., Furuno.M., Hanagaki.T., Hara.A., Hashizume.W.,
   Hayashida.K., Hayatsu.N., Hiramoto.K., Hiraoka.T., Hirozane.T.,
   Hori.F., Imotani.K., Ishii.Y., Itoh.M., Kagawa.I., Kasukawa.T.,
   Kato.H., Kawai.J., Kojima.Y., Kondo.S., Konno.H., Kouda.M.,
   Koya.S., Kurihara.C., Matsuyama.T., Miyazaki.A., Murata.M.,
   Nakamura.M., Nishi.K., Nomura.K., Numazaki.R., Ohno.M., Ohsato.N.,
   Okazaki.Y., Saito.R., Saitoh.H., Sakai.C., Sakai.K., Sakazume.N.,
   Sano.H., Sasaki.D., Shibata.K., Shinagawa.A., Shiraki.T.,
   Sogabe.Y., Tagami.M., Tagawa.A., Takahashi.F., Takaku-Akahira.S.,
   Takeda.Y., Tanaka.T., Tomaru.A., Toya.T., Yasunishi.A.,
   Muramatsu.M. and Hayashizaki.Y.
   Direct Submission
   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
   Physical and Chemical Research (RIKEN), Laboratory for Genome
   Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
   RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
   Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
   URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
   Fax:81-45-503-9216)
7  cDNA library was prepared and sequenced in Mouse Genome
   Encyclopedia Project of Genome Exploration Research Group in Riken
   Genomic Sciences Center and Genome Science Laboratory in RIKEN.
   Division of Experimental Animal Research in Riken contributed to
   prepare mouse tissues.
   Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
   Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
   Trust/MRC building Addenbrookes Hospital Cambridge) whose
   assistance we gratefully acknowledge.
   Please visit our web site for further details.
   URL:http://genome.gsc.riken.go.jp/
   URL:http://fantom.gsc.riken.go.jp/.
   Location/Qualifiers
       1. 1706
          /organism="Mus musculus"
          /mol_type="mRNA"
          /strain="C57BL/6J"
          /db_xref="FANTOM,DB:A630059111"
          /db_xref="MG1:2407136"
          /db_xref="taxon:10090"
          /clone="A630059111"
          /tissue_type="thymus"
          /clone_lib="RIKEN full-length enriched mouse cDNA library"
          /dev_stage="3 days neonate"
          158..1516
          /note="unnamed protein product; Friend leukemia
          integration 1 (MGD|MG1:95554, GB|NM_008026, evidence:
          BLASTN, 99%, match=1671)
8  CDS

```

```

putative"
/codon_start=1
/protein_id="BAC31169.1"
/db_xref="GI:26334937"
/translation="MDGTIKKALSVVSDQSLFSDSAYGAAHLPKADMTASGSPDYQG
PHKINLPPOEWINQPVVVKREYDHNGSRESFVDCSVKCNKLAVGGEANPMNY
NSYMDKNGPPPMNTNERRVTPADPTLTQEHVROWLEWAIKEYGLMEIDTSPFO
NMGKELCVANKEDFIRASAYNTEVLVLSHLSVRESSLAYNTTSHDQSSRLNVE
DPSQIVRRGANNNNSGLKPLGGSTMGKNTQEPQDPYQIOLGPTSSRLANP
GSGCIQWFLLSLSDSANASCTWEGTNGEPMKTDPEVARMGERSKPMNVDK
LSRALYYDKNIMTKVHGKRYAYKDFHGIQAQPHPTETSMYKYPDSISWPSVH
AHQCKVNFVPSHPSPMPTSSSPFGAASQYWTSTAGIYFNPSPVRPHENTHVPFSLHS
YY"
Alignment Scores:
Pred. No.: 8.02e-129 Length: 1706
Score: 1613.50 Matches: 305
Percent Similarity: 78.99% Conservative: 56
Best Local Similarity: 66.74% Mismatches: 85
Query Match: 11 Indels: 11
DB: 7 Gaps: 7
US-09-902-772-2 (1-451) x AK042111 (1-1706)
QY 1 MetAlaSerThrIleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGlu 20
DB 158 ATGCACGGGACTATTAAAGGAGGCTCTCTGTGTGTGTGAGTCAGCATCAGTCCCTTTTGTAT 217
QY 21 CysAlaTyrGly---SerProHisLeuAlaLysThrGluMetThrAlaSerSerSer 39
DB 218 TCAGCATACGGAGCGGCGGCCATCTCCCAAGGCAGATATGATGTTGTCGGGAGTCTCT 277
QY 40 GluTyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGln 59
DB 278 GACTACGGCGGACCCCAAAATCAACCCCTGCGCACCGCAGCAGGAGTGGATCAACCCAG 337
QY 60 ProProAlaArgValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArg 79
DB 338 ---CCAGTGAAGAGTCAATGTCAAGCGGAGTAT-----GACCACATGAATGGATCCAGG 388
QY 80 AsnSerProAspAspCysSerValAlaLysGlyGlyLysMetValSerSerSerAspAsn 99
DB 389 GAGTCTCGGTGGAGTGCAGTGTGAGCAATGTAACAAGTGTGTGGCGGAGGCGGAGGCC 448
QY 100 ValGlyMetAsnTyrGlySerTyrMetGluGluLysHis---11eProProAsnMet 118
DB 449 AACCCCATGAATATATAGTATGATGATGAGAGAACCGGCCCTCTCTCCCAATG 508
QY 119 ThrThrAsnGluArgValIleValProAlaAspProThrLeuTrpSerThrAspHis 138
DB 509 ACCACCAACGACGAGGAGTCAATGTGCTGTCAGACCCACACACTGTGGACAGGAGCAC 568
QY 139 ValArgGlnTrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeu 158
DB 569 GTTCACAGTGGCTGGAGTGGGCTATATAAGGAATACGGATTCATGGAGGAGTTCACACTTCC 628
QY 159 LeuPheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArg 178
DB 629 TTCCTCCAGAACATGATGGCAAGGAATGTGTAATGAACACAGGAGGACTTCTCTCCGA 688
QY 179 LeuThrProSerTyrAsnAlaAspIleLeuSerHisLeuHisTyrLeuArgGluArg 198
DB 689 GCCACCTCCGCTACACACAGAGAGTGTGTGTGCGCACCTCAGTTACCTCAGGGA--- 745
QY 199 GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThr 218
DB 746 ---AGTCTACTGCTGGCTTATAACACCACTCCCAACAGCAGGACTCTCTCAGACTGAAT 802
QY 219 ThrArgProAspLeuProTyrGluGlnAlaArgArgSerAlaTrpThrSerHisSerHis 238
DB 803 GTCAGGAGAGACCTTCTTATGACTCTGTGAGGAGGAGGAGCATGGAACAATAATATGAAC 862

```

QY	239	ProThrGlnSerLysAla---ThrGlnProSerSerSerThrValProLysThrGluAsp	257
Db	863	TCCTGGGCTCAACAAAAGTCTCTCCCTGGAGGATCACAGACCACTGGCGAAGAACACTGAG	922
QY	258	GlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsn	277
Db	923	CAGCGGCCCCAGCCAGATCCTTATCAGATCCTGGGCGCAACAGAGAGCCCTAGCAAC	982
QY	278	ProGlySerGlyGlnIleGlnLeuTyrGlnPheLeuLeuGluLeuLeuSerAspSerSer	297
Db	983	CTCTGGAGTGGGCAGATCCAGCTGTGCGAGTTCTCTGGAACTACTGTCCGACAGCGCC	1042
QY	298	AsnSerAsnCysIleThrTyrGluGlyThrAsnGlyGluPheLysMetThrAspProAsp	317
Db	1043	AAGCCAGCTGTATCACCTGGGAGGGACCAACGGGGAGTTCAAAATGACGGACCCCTGAT	1102
QY	318	GluValAlaArgArgTyrGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeu	337
Db	1103	GAGGTGCCAGGCGCTGGGGAGAGCGAAGCAAGCCCAACATGAATATTATGACAAGCTG	1162
QY	338	SerArgAlaLeuArgTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArg	357
Db	1163	AGCGCGGCCCTCCGACTACTACTATGACAAAACATTTATGACCAAGTGCATGCCAAAGG	1222
QY	358	TyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGlu	377
Db	1223	TATGCCCTACAGTTTGACTTCCATGGCATGGCCAGGCCCTGGAGCCACATCCACAGAG	1282
QY	378	SerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisPro	397
Db	1283	ACATCCATGTACAAAGTATCCCTCTGTATCTCTACATGCTCTTCTACCATGCCATCAA	1342
QY	398	GlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSerSerPhe	417
Db	1343	CAGAAGGTGAACTTGTGCCGTCTCACCCATCTCCATGCTCTGCACCTCTCCAGCTTC	1402
QY	418	PheAlaAlaProAsnProTyrTyrPAsnSerProThrGlyGlyIleTyrProAsn----	435
Db	1403	TTTGGAGCAGCATCACAAATCTGGAGCTCCCCACTGCTGGATCTATCCAAACCCCACT	1462
QY	436	---ThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr	451
Db	1463	GTCCCCCGCCATCCTTAACACCCAGTGCCTTCACACTTAGCAGCTACTACTAC	1513
RESULT 7			
AY418893		1347 bp DNA linear GSS 17-DEC-2003	
LOCUS			
DEFINITION		Homo sapiens FLI1 gene, VIRTUAL TRANSCRIPT, partial sequence,	
ACCESSION		AY418893	
VERSION		AY418893.1	
KEYWORDS		GSS.	
SOURCE		Homo sapiens (human)	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
		1 (bases 1 to 1347)	
		Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejarawal, A.,	
		Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,	
		Perriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,	
		Adams, M.D. and Cargill, M.	
		Inferring nonneutral evolution from human-chimp-mouse orthologous	
		gene trios	
JOURNAL		Science 302 (5652), 1960-1963 (2003)	
PUBMED		14671302	
REFERENCE		2 (bases 1 to 1347)	
AUTHORS		Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejarawal, A.,	
		Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,	
		Perriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,	
		Adams, M.D. and Cargill, M.	
		Direct Submission	
TITLE		Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,	
JOURNAL		Rockville, MD 20850, USA	

COMMENT These sequences were made by sequencing genomic exons and ordering them based on alignment.

FEATURES  
source Location/Qualifiers

```

1..1347
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
<1..>1347
    /gene="FLI1"
    /locus_tag="HCM6709"
```

ORIGIN

Alignment Scores:

Pred. No.:	1.01e-128	Length:	1347
Score:	1610.50	Matches:	302
Percent Similarity:	79.69%	Conservative:	59
Best Local Similarity:	66.67%	Mismatches:	81
Query Match:	65.90%	Indels:	11
DB:	29	Gaps:	7

US-09-902-772-2 (1-451) x AY418893 (1-1347)

```

QY      5   IleLysGluAlaLeuSerValValSerGlnAspGlnSerLeuPheGluCysAlaTyrGly  24
       :::::::::::::::::::::
Db      1   CTGCAGGAGGCTCTGCGTGGTGAGGACGCCAGGATCCCTTTGACTCAGCGTCAGGA  60
       :::::::::::::::::::::

QY     25   ---SerProHisLeuAlaLysThrGluMetThrAlaSerSerSerSerGluTyrGlyGln  43
       :::::::::::::::::::::
Db     61   GGGCGAGGCCATCTCCCCAAGCGCATGACTCCCTCGGGGAGTCTCTGACTACGGGCAG  120
       :::::::::::::::::::::

QY     44   ThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnProProAlaArg  63
       :::::::::::::::::::::
Db    121   CCCCAAGAATCAACCCCCTCCACCACAGCAGGAGTGGATCAATCAG---CCAGTCAGG  177
       :::::::::::::::::::::

QY     64   ValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSerProAsp  83
       :::::::::::::::::::::
Db    178   GTCAACGTCGAACGGGAGTAT-----GACCACATGAATGGATCCAGGGAGTCTCCGGT  231
       :::::::::::::::::::::

QY     84   AspCysSerValAlaLysGlyGlyLysMetValSerSerSerSerAspAsnValGlyMetAsn  103
       :::::::::::::::::::::
Db    232   GACTGCAGCGTTAGCAATGCAGCAAGCTGGTGGCGGAGCGAGTCCAACCCCATGAAC  291
       :::::::::::::::::::::

QY    104   TyrGlySerTyMetGluGluLysHis----IleProProAsnMetThrThrAsnGlu  122
       :::::::::::::::::::::
Db    292   TACAACAGCTATATGGACGAGAAGATGGCCCCCTCTCTCCAAATGACCAACCAACGAG  351
       :::::::::::::::::::::

QY    123   ArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGlnTrp  142
       :::::::::::::::::::::
Db    352   AGGAGAGTCATCGTCCCGCGAGACCCCACTGTGGACACAGGAGCATGTGAGGCATGG  411
       :::::::::::::::::::::

QY    143   LeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIleLeuLeuPheGlnAsn  162
       :::::::::::::::::::::
Db    412   CTGGAGTGGGCCATAAAGGAGTACAGCTTGTATGGAGATCGACACATCCTTTTTCCAGAAC  471
       :::::::::::::::::::::

QY    163   IleAspGlyLysGluLeuCysLysMetThrylSaspAspPheGlnArgLeuThrProSer  182
       :::::::::::::::::::::
Db    472   ATGGATGGCAAGGARCTGTGTAAATGAACAAGGAGGACTTCCTCCGCGCACCCCTC  531
       :::::::::::::::::::::

QY    183   TyrAsnAlaAspIleLeuSerHisLeuHisTyrlLeuArgGluArgGlyAlaThrPhe  202
       :::::::::::::::::::::
Db    532   TACAACACCGAAGTGTGTGTACACCTCAGTTACCTCAGGGAA-----AGTTCACGT  581
       :::::::::::::::::::::

QY    203   IlePheProAsnThrSerValTyProGluAlaThrGlnArgIleThrThrArgProAsp  222
       :::::::::::::::::::::
Db    586   CTGGCCCTATAATACAACTCCACACCCAGCAACCACTCCATCGATTGAGTGTCAAAGAAGAC  645
       :::::::::::::::::::::

QY    223   LeuProTyrgluGlnAlaArgArgSerAlaTrpThrSerHisSerHisProThrGlnSer  244
       :::::::::::::::::::::
Db    646   CCTTCTTATGATCTAGTCNAGAGAGGAGCTTGGGGCAATAACATGAATTCGTGGCCCTCAAC  705
       :::::::::::::::::::::

QY    243   LysAlaThrGln---ProSerSerSerThrValProLysThrGluAspGlnArgProGln  266
       :::::::::::::::::::::
Db    706   AAAAGTCTCTCCCTTGGAGGGGCAAAACGATCAGTAGAATACAGACGACGCGCCCGAC  765
       :::::::::::::::::::::
```

Qy	262	LeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGlySerDly	281
Db	766	CCAGATCCGATCAGATCTCTGGGCCCGACAGCAGTCGCTAGCCAACTCGAAGCGGG	825
Qy	282	GlnIleGlnLeuTyrGlnPheLeuLeuGluLeuSerAspSerSerAsnSerAsnCys	301
Db	826	CAGATCCAGCTGTGGCAATCTCTCTGGAGTCTCTCCGACAGCCCAAGCCAGCTGT	885
Qy	302	IleThrTyrGluGlyThrAsnGlyGluPheLysMetThrAspProAspGluValAlaArg	321
Db	886	ATCACCTGGAGGGAGCAACCGGGAGTTCAAAATGACGACCCCGATGAGGTGGCCAGG	945
Qy	322	ArgTyrGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArgAlaLeu	341
Db	946	CGTGGGGCGAGGGAACCAAGCAACCAATGATGACCAAGCTGAGCGGGCCCTC	1005
Qy	342	ArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaTyrLys	361
Db	1006	CGTATTACTATGATAAAACATTATGACCAAGTGCACGGCAAAAGATATGTTACAAA	1065
Qy	362	PheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGluSerSerMetTyr	381
Db	1066	TTTGACTTCCAGCGATTCGCCAGGCTCTCCAGCCATCCGACGAGTGTCCAGTAC	1125
Qy	382	LysTyrProSerAspLeuProTyrMetSerTyrHisAlaHisProGlnLysMetAsn	401
Db	1126	AAGTACCCTTCTGACATCTCTACATGCTCTCTACCATGCCACCAGCAGAGGTGAAC	1185
Qy	402	PheValAlaProHisProProAlaLeuProValThrSerSerPheAlaAlaPro	421
Db	1186	TTTGTCTCTCCCATCATCTCTCCAGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCT	1245
Qy	422	AsnProTyrTyrAsnSerProThrGlyGlyIleTyrProAsn	438
Db	1246	TCACAACTAGGACTCCCGACGCGGGGGAATCTACCCCAACCCCAACGTCGCCCAT	1305
Qy	439	ProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr	451
Db	1306	CCTAACACCCACGTCCTTCCACTTAGCGAGCTACTAC	1344
RESULT 8			
LOCUS	AK036655	3118 bp mRNA linear HTC 19-SEP-2003	
DEFINITION		Mus musculus adult male bone cDNA, RIKEN full-length enriched library, clone:9830146N17 product:Friend leukemia integration 1, full insert sequence.	
ACCESSION	AK036655		
VERSION	AK036655.1	GI:26331575	
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			

Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000)

20530913

11076861

4

The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

Functional annotation of a full-length mouse cDNA collection. Nature 409, 685-690 (2001)

5

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002)

6 (bases 1 to 3118)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M., Hayashida, K., Hayateu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

Direct Submission

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.go.jp/

URL: http://fantom.gsc.riken.go.jp/

Location/Qualifiers

1. 3118

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="PANTOM\_DB:9830146N17"

/db\_xref="MGI:2401092"

/db\_xref="taxon:10090"

/clone="9830146N17"

/sex="male"

/tissue\_type="bone"

/clone\_lib="RIKEN full-length enriched mouse cDNA library"

/dev\_stage="adult"

192. 1547

/note="unnamed protein product; Friend leukemia integration 1 (MGD|MGI:95554, GB|NM\_008026, evidence: BLASTN, 99%, match=1671) putative"

/codon\_start=1

/protein\_id="BAC29518.1"

/db\_xref="GI:26331576"

/translation="MDGTFKALSVSDQSLFDSAYGAARHLPKADMTASGSPDYGQ PEKINLPQQEWINQPVVNVKRGSDSPVDCVCKNKLVGCGANPNYN SYMDKNGPPPNMTNRRVTFADPLTQEHVQRLWEALTEIDTSFFQN MDGKELCKNKKEDFLRATSAVNTVLLSHLSYRESSLLAYNTTTSHTQSSRLNKED



PSYDVRGAWNNMNSGLKPLGSGSTWKNTEORPODPVQILGPTSSRLANPG  
SGOIQWQLELLSDSANASCIYTWETNGSEFKWTDDEVARSWGSEKSKPNVNYDKL  
SRALYVYKXNIMTKVGYKAYKDPHGIAQALQPHPTETSMYKVPDSISYMPSHA  
HQQKVNVPFSPNPSSNPVTSISFFGAASQYIWTISFTAGIYPNPSVPRPNTHVPSHLSGY  
Y"

ORIGIN

Alignment Scores:  
Pred. No.: 2.01e-127 Length: 3118  
Score: 1602.00 Matches: 305  
Percent Similarity: 78.77% Conservative: 55  
Best Local Similarity: 66.74% Mismatches: 85  
Query Match: 65.55% Indels: 12  
DB: 11 Gaps: 8

US-09-902-772-2 (1-451) x AK036655 (1-3118)

QY 1 MetAlaSerThrIleLeuValSerGluAspGlnSerLeuPheGlu 20  
Db 192 ATGGACGGGACTATTAAAGGAGGCTCTGTCTGTGTGAGTGACGATCAGTCCCTTTTGTAT 251  
QY 21 CysAlaTyrGly---SerProHisLeuAlaLysThrGluMetThrAlaSerSerSer 39  
Db 252 TCAGCATACGAGCGGCGGAGCCCATCTCCCAAGGCGAGTATGACTGCTTCGGGAGTCCT 311  
QY 40 GluTyrGlyGlnThrSerLysMetSerProArgValProGlnGlnAspTyrLeuSerGln 59  
Db 312 GACTACGGGCGGACCCCAAAATCAACCCCTGCGCCAGGAGGAGTGGATCAACAG 371  
QY 60 ProProAlaArgValThrIleLysMetGluCysAsnProAlaGlnValAsnGlySerArg 79  
Db 372 ---CCAGTCAGAGTCATGTCAAGCGGAGTAT-----GACCACATGATGATCCAGG 422  
QY 80 AsnSerProAspAspCysSerValAlaLysGlyGlyLysMetValSerSerSerAspAsn 99  
Db 423 GAGTCTCCGGTGGAGTCAGTGTGTC---AAATGTAACAGTGTGTGGCGGAGGCGAAGCC 479  
QY 100 ValGlyMetAsnTyrGlySerTyrMetCluGluLysHis----IleProProAsnMet 118  
Db 480 AACCCCATGAACTAATAGTCTATGATGATGAGAGGAGGCGCCCTCTCTCCCAACATG 539  
QY 119 ThrThrAsnGluArgValIleValProAlaAspProThrLeuTyrSerThrAspHis 138  
Db 540 ACCACCAACGAGGAGTGTGCTGCTGACAGCCCACTGTGGACACAGGAGCAC 599  
QY 139 ValArgGlnTyrLeuGluTyrAlaValLysGluTyrGlyLeuProAspValAspIleLeu 158  
Db 600 GTTCACAGTGGCTGGAGTGGCTATAAAGGAATACGGATTTGATGGAGATGACACTTCC 659  
QY 159 LeuPheGlnAsnIleAspGlyLysGluLeuLysMetThrLysAspAspPheGlnArg 178  
Db 660 TTCITCCAGACATGATGGCAAGGAATGTGTAATGAACAAGGAGGACTTCTCCCGA 719  
QY 179 LeuThrProSerTyrAsnAlaAspIleLeuLysHisLeuHisTyrLeuArgGluArg 198  
Db 720 GCCACCTCGGCTACACAGAAAGTGTGTGTGCGCACCTCAGTACCTCAGGAA--- 776  
QY 199 GlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIleThr 218  
Db 777 ---AGTTCAGTCTGGCTTAAACACACCTCCCATACAGACCACTGCTCAGCTGAAT 833  
QY 219 ThrArgProAspLeuProTyrGluGlnAlaArgSerAlaTyrThrSerHisSerHis 238  
Db 834 GTCAGGAGAGCCCTTCCTATGACTCTGTACGAGGAGGAGGATGGAACAATAATATGAAC 893  
QY 239 ProThrGlnSerLysAla---ThrGlnProSerSerThrValProLysThrGluAsp 257  
Db 894 TGTGCGCTCAACAAAGTCCTCTCTGAGGATCACAGACCATGGCGGCAAGAACACTGAG 953  
QY 258 GlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsn 277  
Db 954 CAGCGGCCCCAGCCAGATCCTTATCAGATCTCTGGGGCCAAACAGCAGCGCGCTAGCAAC 1013

QY 278 ProGlySerGlyGlnIleGlnLeuTyrGlnPheLeuLeuGluLeuLeuSerAspSerSer 297  
Db 1014 CTGGGAGTGGGACAGTCCAGCTGTGGCAGTTTCTCTGGAACCTACTGTCCGACAGCGCC 1073  
QY 298 AsnSerAsnCysIleThrTyrGluGlyThrAsnGlyGluPheLysMetThrAspProAsp 317  
Db 1074 AACGCCAGCTGTATCACCTGGGAGGGGACCAACGGGGAGTTCAAAATGACGGACCTGAT 1133  
QY 318 GluValAlaArgArgTyrGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeu 337  
Db 1134 GAGTGGCCAGGCGCTGGGAGAGCGGAGAGCAACCAACATCAATATATCACAAGCTG 1193  
QY 338 SerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArg 357  
Db 1194 AGCCGGGCGCTCCGATACCTATGACAAAACATATGACCAAAAGTGCATGCCAAAGG 1253  
QY 358 TyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGlu 377  
Db 1254 TATGCTACAGTTTGACTTCCATGGCATTTGCCAGGCGCTGCAGCCACATCCACAGAG 1313  
QY 378 SerSerMetTyrLysTyrProSerAspLeuProTyrMetSerTyrThrHisAlaHisPro 397  
Db 1314 ACATCCATGTACAAGTATCCCTCTGATATCTCTACATGCTCTCTACCATGCCATCAA 1373  
QY 398 GlnLysMetAsnPheValAlaProHisProAlaLeuProValThrSerSerSerPhe 417  
Db 1374 CAGAAGTGAATTTGTCCCGTCTCACCATCTCCATGCTGTCCCTCTCCAGCTTC 1433  
QY 418 PheAlaAlaProAsnProTyrTyrAsnSerProThrGlyGlyIleTyrProAsn----- 435  
Db 1434 TTTGAGCAGCATCAATACCTGACCTCCCTCCCTGCTGCTGCTATCCAAACCCAGT 1493  
QY 436 ---ThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451  
Db 1494 GTCCCCCGGCATCTTAACACCCAGCTGCTTCCACCTAGGCGAGCTACTAC 1544

RESULT 9  
AY418895 1347 bp DNA linear GSS 17-DEC-2003  
LOCUS Mus musculus FLII gene, VIRTUAL TRANSCRIPT, partial sequence,  
DEFINITION genomic survey sequence.  
ACCESSION AY418895  
VERSION AY418895.1 GI:39774855  
KEYWORDS GSS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 1347)  
AUTHORS Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,  
Todd, M.A., Tanenbaum, D.M., Civallo, D.R., Lu, F., Murphy, B.,  
Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,  
Adams, M.D. and Cargill, M.  
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous  
gene trios  
JOURNAL Science 302 (5652), 1960-1963 (2003)  
PUBMED 14671302  
REFERENCE 2 (bases 1 to 1347)  
AUTHORS Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,  
Todd, M.A., Tanenbaum, D.M., Civallo, D.R., Lu, F., Murphy, B.,  
Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,  
Adams, M.D. and Cargill, M.  
TITLE Direct Submission  
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
Rockville, MD 20850, USA  
COMMENT These sequences were made by sequencing genomic exons and ordering  
them based on alignment  
FEATURES  
source 1..1347  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
<1..->1347  
gene



/gene="FLI1"		/locus_tag="HCM6709"	
ORIGIN			
Alignment Scores:			
Pred. No.:	9.1e-128	Length:	1347
Score:	1599.50	Matches:	301
Percent Similarity:	79.25%	Conservative:	58
Best Local Similarity:	66.45%	Mismatches:	83
Query Match:	65.45%	Indels:	11
DB:	29	Gaps:	7
US-09-902-772-2 (1-451) x AY418895 (1-1347)			
QY	5	IleLysGluAlaLeuSerValValSerGluAspGlnSerLeuPheGluCysAlaTyGly	24
DB	1	TTTCAGGAGGCTCTCTGTGTGAGTACGATCAGTCCCTTTTGATCAGCATACGA	60
QY	25	---SerProHisLeuAlaLysThrGluMetThrAlaSerSerSerSerGluTyGln	43
DB	61	GCGGAGCCCATCTCCCAAGGAGATGACTGCTTCGGGAGTCTGTACTACGGGCG	120
QY	44	ThrSerLysMetSerProArgValProGlnGlnAspTrpLeuSerGlnProProAlaArg	63
DB	121	CCCCACAAATCAACCCCTGCCACCGCAGGAGTGGATCAACACAG---CCAGTGAGA	177
QY	64	ValThrIleLysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSerProAsp	83
DB	178	GTCATGTCAAGCGGAGTAT-----GACCACATGATGATCCAGGAGTCTCCGGTG	231
QY	84	AspCysSerValAlaLysGlyGlyLysMetValSerSerSerSerAspAsnValGlyMetAsn	103
DB	232	GACTGCAGTGTCAAGCAATGTAAACAGCTGTGTGGCGGAGCGCAAGCAACCCCATGAAC	291
QY	104	TyrGlySerTyrMetGluGluLysHis---IleProProProAsnMetThrThrAsnGlu	122
DB	292	TATATAGTGTACATGATGAGAGAACGGCCCCCTCTCCACATGACACCAACGAA	351
QY	123	ArgArgValIleValProAlaAspProThrLeuTrpSerThrAspHisValArgGlnTrp	142
DB	352	CGGAGAGTCAATGTGCTGCAGACCCACACATGTGGACACAGGACAGTTCGACAGTGG	411
QY	143	LeuGluTrpAlaValLysGluTyGlyLeuProAspValAspIleLeuLeuPheGlnAsn	162
DB	412	CTGGATGGGCTATAAAGAAATACGATTGATGAGATTGACATCTCTTCTTCCAGAAC	471
QY	163	IleAspGlyLysGluLeuCysLysMetThrLysAspAspPheGlnArgLeuThrProSer	182
DB	472	ATGGATGGCAAGGAATGTGTATAAATGAACAAGGAGGACTTCTCCGAGCCACCTCCGCC	531
QY	183	TyrAsnAlaAspIleLeuSerHisLeuHisTyLeuArgGluArgGlyAlaThrPhe	202
DB	532	TACACACAGAAAGTGTGTGTGGACCTCAGTTACCTACCGGAA-----AGTTCAGTG	585
QY	203	IlePheProAsnThrSerValTyProGluAlaThrGlnArgIleThrArgProAsp	222
DB	586	CTGGCTTATAACACACCTCCATACAGACCCAGTCTCTCAGCTGAATGTCAAGGAAGAC	645
QY	223	LeuProTyGluGlnAlaArgSerAlaTrpThrSerHisSerHisProThrGlnSer	242
DB	646	CTTCTTATGACTCTCTCAGGAGAGGAGATGAACAATAATATGAATCTGGCCTCAAC	705
QY	243	LysAla---ThrGlnProSerSerThrValProLysThrGluAspGlnArgProGln	261
DB	706	AAAGTCTCTCTCTGGAGATCAGACCATCGGCAAGAACACTGAGCAGCGGCCCGAG	765
QY	262	LeuAspProTyGlnIleLeuGlyProThrSerSerArgLeuAlaAsnProGlySerGly	281
DB	766	CCAGATCTCTTATCAGATCTCTGGGGCCCAACACGACCGCTAGCAAAACCTTGGGAGTGG	825
QY	282	GlnIleGlnLeuTrpGlnPheLeuLeuGluLeuLeuSerAspSerSerAsnSerAsnCys	301
DB	826	CAGATCCAGTGTGGCGATTTCTCTCTGGAACTACTGTCCGACAGCGCCACCGCAGCTGT	885

Alignment Scores:			
Pred. No.:	5,898-113	Length:	1256
Score:	1428.00	Matches:	270
Percent Similarity:	75.83%	Conservative:	50
Best Local Similarity:	63.98%	Mismatches:	92
Query Match:	29	Indels:	10
DB:	28	Gaps:	6
US-09-902-772-2 (1-451) x AY418894 (1-1256)			
Qy	35	AlaSerSerSerGluThrGlnThrSerLysMetSerProAsgValProGlnGln	54
Db	3	GCCTCGGGAGTCGCGAGTCCGAGCCGACAGATCAACCCCTCCACCCANNNG	62
Qy	55	AspTTPLeuSerGlnProProAlaArgValThrIleLysMetGluCysAsnProAsnGln	74
Db	63	GAGTGGATCAANN--NCNNNAGGTCAACNNNAAGCGG-----GNNNATGACAC	113
Qy	75	VallanGlySerArgAsnSerProAspAspCysSerValAlaLysGlyGlyMetVal	94
Db	114	ATGAATGGATCCAGGAGTCTCCGGTGGACTCGACGCTTACGAAATGCGAAGCTGGT	173
Qy	95	SerSerSerAspAsnValGlyMetAsnTyrGlySerTyrMetGluGluLysHis--Ile	113
Db	174	GCGGAGCGAGTCCACCCCATGACTACACAGTATATGAGGAGAGATGGCCCC	233
Qy	114	ProProProAsnMetThrThrAsnGluArgArgValIleValProAlaAspProThrLeu	133
Db	234	CCTCTCCCAACATGACCAACAGAGAGAGTCTATCGTCCCGGAGAGCCACACTG	293
Qy	134	TrpSerThrAspHisValArgGluTrpLeuGluTrpAlaValLysGluTyrGlyLeuPro	153
Db	294	TGGACACAGGAGTGTGAGGCAATGGTGGAGTGGCCATTAAGAGGTACAGTTGATG	353
Qy	154	AspValAspIleLeuLeuPheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLys	173
Db	354	GAGATCGACATCCTTTTCCAGAACATGATGATGCAAGGAACTGTGTAAATGAACAAG	413
Qy	174	AspAspPheGlnArgLeuThrProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHis	193
Db	414	GAGGATCTTCCCGTGGCCACACCTCTACACAGCGAAGTGTGTGTACACCTCAGT	473
Qy	194	TyrLeuArgGluArgGlyValAlaPheIlePheProAsnThrSerValTyrProGluAla	213
Db	474	TACCTCAGGAA-----AGTCACTGTGGCCATATATACACCTCCACAGCACCA	527
Qy	214	ThrGlnArgIleThrThrArgProAspLeuProTyrGluGlnAlaArgArgSerAlaTrp	233
Db	528	TCTCAGCACTGAGTGTCAAGAGAGACCTTCTTATGACTCAGTCAGAAGAGGAGCATGG	587
Qy	234	ThrSerHisSerHisProThrGlnSerLysAlaThrGln--ProSerSerSerThrVal	252
Db	588	GGCAATACATGAATCTGGCTCAACAAAGTCTCTCCCTGGAGGGGCAACACGATC	647
Qy	253	ProLysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSer	272
Db	648	AGTAAGAATACAGACCAACGCGCCAGCAGATCGTATCAGATCTGGGCGCCAGCAGC	707
Qy	273	SerArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeu	292
Db	708	AGTCCCTCAGNNNNNAGAGCGGCGAGATCCAGCTGTGGCAATCTCTCTGGAGCTG	767
Qy	293	LeuSerAspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLys	312
Db	768	CTCTCCGAGCGGCAACGCGAGTGTATCCTCTGGAGGGGAGCAACCGGGAGTTCAA	827
Qy	313	MetThrAspProAspGluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMet	332
Db	828	ATGAGCGGACCGATGAGTGGCCAGGCGCTGGGCGAGCGGAAAGCAAGCCCAACATG	887
Qy	333	AsnTyrAspLysLeuSerArgAlaLeuArgTyrTyrAspLysAsnIleMetThrLys	352

Db	888	AATTACACAGCTGAGCGCGCCCTCGTTATTACTATGATAAAACATTATGACAAA	947
Qy	353	ValHisGlyLysArgTyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGln	372
Db	948	GTGCACGGCAAAAGATATGCTTACAAATTGACTTCCACGGCATTCGCCAGGCTCTGCAG	1007
Qy	373	ProHisProProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSer	392
Db	1008	CCACATCCGACGAGTCGTCATGACAAAGTACCCCTCTGACATCTCTACATGCTTCC	1067
Qy	393	TyrHisAlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProVal	412
Db	1068	TACCATGCCACACAGAGAGTGAATTTGCTCCCTCCCATCCATCTCTCCATGCTGTC	1127
Qy	413	ThrSerSerSerPhePheAlaAlaProAsnProTyrTyrAsnSerProThrGlyGlyIle	432
Db	1128	ACTTCTCCAGTCTCTTTGGAGCGGCATCAATACTGACCTCCGCCANNNGGGAATC	1187
Qy	433	TyrProAsn-----ThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThr	449
Db	1188	TACCCCAACCCCAACGTCCTCCCGCCATCTTAACACCAAGTGGCTTCACATAGGCAGC	1247
Qy	450	TyrTyr 451	
Db	1248	TACTAC 1253	
RESULT 11			
BU334592			
LOCUS			
DEFINITION			
sequence.			
ACCESSION			
BU334592			
VERSION			
BU334592.1			
KEYWORDS			
EST.			
SOURCE			
Gallus gallus (chicken)			
ORGANISM			
Gallus gallus			
REFERENCE			
1 (bases 1 to 726)			
AUTHORS			
Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,			
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.			
TITLE			
A Comprehensive Collection of Chicken cDNAs			
JOURNAL			
Curt. Biol. 12 (22), 1965-1969 (2002)			
MEDLINE			
22335534			
PUBMED			
12445392			
COMMENT			
Contact: Simon Hubbard			
Department of Biomolecular Sciences			
University of Manchester Institute of Science and Technology			
(UMIST)			
PO Box 88, Manchester, M60 1QD, UK			
Tel: 01612008930			
Fax: 01612360409			
Email: Simon.Hubbard@umist.ac.uk.			
Location/Qualifiers			
1. .726			
/organism="Gallus gallus"			
/mol_type="mRNA"			
/strain="White Leghorn, Hisex"			
/db_xref="taxon:9031"			
/clone="ChEST418m23"			
/tissue_type="whole embryo"			
/dev_stage="10"			
/lab_host="DH10B"			
/notes="Organ: whole embryo; Vector: pBluescript II KS(+);			
Site 1: EcoRI; Site 2: NotI; This normalized library was			
constructed from 1 million independent clones. cDNA			
synthesis was initiated using an oligo(dT) primer, using			
methylated C in the first strand synthesis reaction.			
Following this first strand reaction, double-stranded cDNA			
was blunted, ligated to NotI adapters, digested with			
EcoRI, size-selected, and cloned into the NotI and EcoRI			

FEATURES

source	
1. .726	
/organism="Gallus gallus"	
/mol_type="mRNA"	
/strain="White Leghorn, Hisex"	
/db_xref="taxon:9031"	
/clone="ChEST418m23"	
/tissue_type="whole embryo"	
/dev_stage="10"	
/lab_host="DH10B"	
/notes="Organ: whole embryo; Vector: pBluescript II KS(+);	
Site 1: EcoRI; Site 2: NotI; This normalized library was	
constructed from 1 million independent clones. cDNA	
synthesis was initiated using an oligo(dT) primer, using	
methylated C in the first strand synthesis reaction.	
Following this first strand reaction, double-stranded cDNA	
was blunted, ligated to NotI adapters, digested with	
EcoRI, size-selected, and cloned into the NotI and EcoRI	

compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

## ORIGIN

## Alignment Scores:

Pred. No.: 4,55e-100 Length: 726  
Score: 1275.00 Matches: 236  
Percent Similarity: 99.16% Conservative: 1  
Best Local Similarity: 98.74% Mismatches: 0  
Query Match: 52.17% Indels: 2  
DB: 13 Gaps: 0

US-09-902-772-2 (1-451) x BU334592 (1-726)

Qy 215 GlnArglleThrThrArgProAspLeuProTyGluGlnAlaArgSerAlaTrpThr 234  
Db 6 GAGAGAATAACAACAGCCAGATTTTACCTATGAGCAAGCAGGAGATCAGCGTGACG 67  
Qy 235 SerHisSerHisProThrGlnSerLysAlaThrGlnProSerSerSerThrValProLys 254  
Db 68 AGTCACAGCCATCCCATCTCAGTCAAAGCTACCCACCATCATCTTCAACAGTGCCCAA 127  
Qy 255 ThrGluAspGlnArgProGlnLeuAspProTyGlnlleLeuGlyProThrSerSerArg 274  
Db 128 ACAGAGACGAGCTCCTCAGTTAGATCCTTATCAGATTTCTTGACCGCAGCAGCGCT 187  
Qy 275 LeuAlaAsnProGlySerGlyGlnlleGlnLeuTrpGlnPheLeuLeuGluLeuSer 294  
Db 188 CTTCCAAATCAGGAGGTGGCAGATACAGTATGGCAGTTCCTACGTGAGCTTCTGTGC 247  
Qy 295 AspSerSerAsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThr 314  
Db 248 GACAGCTCCAACTCCAACTGCATCACCTGGAGGGCCAAATGGGAGTTCAAGATGACA 307  
Qy 315 AspProAspGluValAlaArgArgTrpGlyGluArgLysSerLysProAsnMetAsnTyr 334  
Db 308 GACCCTGATGAAGTGGCTCGCGTGTGGGAGAGGAAAGCAACCTAATCATGAACTAT 367  
Qy 335 AspLysLeuSerArgAlaLeuArgTyrTyrTyrAspLysAsnIleMetThrLysValHis 354  
Db 368 GACAAACTCAGCGCGCATTCGCTACTATGACAAATAATATTATGACTAAAGTTTCA 427  
Qy 355 GlyLysArgTyrAlaTyrLysPheAspPheHisGlylleAlaGlnAlaLeuGlnProHis 374  
Db 428 GGTAAACGCTATGCTACAAATTTGATTTCCACGGAATCGCTCAGGCCCTCCAGCCTCAC 487  
Qy 375 ProProGluSerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHis 394  
Db 488 CTTCCAGNATCATTCATGTACAAATACCCATCAGACCTCCCTACATGAGTTCTACCAT 547  
Qy 395 AlaHisProGlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSer 414  
Db 548 GCACACCCCAAGAGATGAATTTGTAGTCTCCCATCCCTGCTTTGGCCCGTAACCTCA 607  
Qy 415 SerSerPhePheAlaAlaProAsnProTyrTrpAsnSerProThrGlyGlylleTyrPr 434  
Db 608 TCCAGCTTTTGTGCCCCCTTAATCCATACGGAATTCACCACTGGGAGCATCTACCC 667  
Qy 434 cAsnThrArgLeuProAlaAlaHisMetProSerHisLeuGlyThrTyrTyr 451  
Db 668 CAATACCAAGGTGCCAGTGTCTATATGCCCTTCCCATCTTTGGCACCTACTAC 720

## RESULT 12

CB269538 709 bp mRNA linear EST 20-FEB-2003  
LOCUS 1008445 Human Fat Cell 5'-Stretch Plus cDNA Library Homo sapiens  
DEFINITION cDNA 5', mRNA sequence.  
ACCESSION CB269538  
VERSION CB269538.1 GI:28444123

## KEYWORDS

SOURCE EST. Homo sapiens (human)

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 709)  
Yang, R.-Z., Shuldiner, A. and Gong, D.-W.

## AUTHORS

EST analysis of human adipose gene expression

## TITLE

## JOURNAL

## COMMENT

Contact: Gong Da-Wei  
Division of Endocrinology, Diabetes and Nutrition

University of Maryland

660 Redwood St, HH497, Baltimore, MD 21201, USA

Tel: 410 706 1672

Fax: 410 706 1622

Email: dgong@medicine.umaryland.edu

## PCR Primers

FORWARD: CTCGGAAGCGCCATTGTGTGGT

BACKWARD: AATACGACTCATTAGGCGGAATTGG

## Seq primer:

GTGTGATCCCGGAATTC.

## FEATURES

## source

1..709

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/sex="Male and Female"

/tissue\_type="Adipose"

/clone\_lib="Human Fat Cell 5'-Stretch Plus cDNA Library"

/note="Vector: lambda triplex"

## ORIGIN

## Alignment Scores:

Pred. No.: 5,84e-87 Length: 709  
Score: 1123.50 Matches: 218  
Percent Similarity: 94.02% Conservative: 2  
Best Local Similarity: 93.16% Mismatches: 13  
Query Match: 45.97% Indels: 4  
DB: 14 Gaps: 1

US-09-902-772-2 (1-451) x CB269538 (1-709)

Qy 204 PheProAsnThrSerValTyrProGluAlaThrGlnArgIleThrThrArgProAsnLeu 223  
Db 11 TTTCNNAATACCTTCAGTATATCTCTGAGCTACGCAAGAATACAACTAGGCCAGATT 70  
Qy 224 ProTyrGluGlnAlaArgSerAlaTrpThrSerHisSerHisProThr---GlnSer 242  
Db 71 CCATATGAGCCCCCAGAGAGATCAGCTGGACCGCTCAGGCCACCCACGCCAGTCG 130  
Qy 243 LysAlaThrGlnProSerSerSerThrValProLysThrGluAspGlnArgProGlnLeu 262  
Db 131 AAAGCTGTCTCAACCATCTCTTCCACAGTGGCCAAACTGAAGACGACGCTCCTCAGTTA 190  
Qy 263 AspProTyrGlnlleLeuGlyProThrSerArgLeuAlaAsnProGlySerGlyGln 282  
Db 191 GATCCTTATCAGATTCTTGGACCAACAGTAGCAGCTTGCAATCCAGGAGTGCCAG 250  
Qy 283 IleGlnLeuTrpGlnPheLeuLeuGluLeuSerAspSerSerAsnSerAsnCysIle 302  
Db 251 ATCCAGCTTTGCGAGTTCTCTCTGAGCTCTCTGCGACAGCTCCCACTCCAGCTGCATC 310  
Qy 303 ThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspProAspGluValAlaArgArg 322  
Db 311 ACCTGGGAAGTCAACAAACGGGAGTTCAAGATGACGGATCCCGACGAGTGCCCGGCGC 370  
Qy 323 TrpGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeuSerArgAlaLeuArg 342  
Db 371 TGGGGAGAGCGGAGAGCAACCACTAGTAACTAGCTCAGCGCGCCCTCCGT 430  
Qy 343 TyrTyrTyrAspLysAsnIleMetThrLysValHisGlyLysArgTyrAlaTyrLysPhe 362  
Db 431 TACTACTGACAGAACATCATGACCAAGGTCCATGGGAAGCGCTACGCCCTACAGTTTC 490

```

QY 363 AspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGluSerSerMetTyrLys 382
Db 491 GACTTCCACGGGATCGCCAGGCCCTCCAGCCACCACCCCGGAGTCATCTCTGACAG 550
QY 383 TyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisProGlnLysMetAsnPhe 402
Db 551 TACCCCTCAGACCT-CCGTACATGGGCTCCTATCAGCCCAACCCACAGAGATGAACCTT 609
QY 403 ValAlaProHisProAlaLeuProValThrSerSerSerPheAlaAlaProAsn 422
Db 610 GTGGCGCCCAACCTCCAGCCCT-CCGTGACATCTTCAGTATTTTTCGTGC-CCAAAC 667
QY 423 ProTyrTrpAsnSerProThrGlyGlyIleTyrProAsnThr 436
Db 668 CCATCTGGATTCACCAACTTGGGGATATACCCCACT 709

RESULT 13
AJ456498 927 bp mRNA linear EST 22-APR-2002
DEFINITION AJ456498 riken1 Gallus gallus cdna clone 8m13r2, mRNA sequence.
ACCESSION AJ456498
VERSION AJ456498.1 GI:20266594
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 827)
Buerstedde,J.M.
Gallus gallus bursal lymphocyte EST
Unpublished (2002)
Contact: Buerstedde JM
Cellular Immunology
Heinrich-Pette-Institute
Martinistr. 52, 20251 Hamburg, Germany
Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
FEATURES
source
1..827
Location/Qualifiers
/organism="Gallus gallus"
/mol_type="mRNA"
/db_xref="taxon:9031"
/clone="8m13r2"
/cell_type="bursal lymphocyte"
/dev_stages="2-3 weeks old"
/clone_lib="riken1"
/note="TCB inbred strain"

ORIGIN
Alignment Scores:
Pred. No.: 2,01e-86 Length: 827
Score: 1118.50 Matches: 211
Percent Similarity: 78.31% Conservative: 20
Best Local Similarity: 71.53% Mismatches: 43
Query Match: 45.77% Indels: 21
DB: 9 Gaps: 3

US-09-902-772-2 (1-451) x AJ456498 (1-827)
QY 138 HisValArgGlnTrpLeuGluTrpAlaValLysGluTyrGlyLeuProAspValAspIle 157
Db 5 CACGTGCCCGCAGTGGCTGGATGGCCATAAAGGATGATGATTAAATGGAGATTGACACC 64
QY 158 LeuLeuPheGlnAsnIleAspGlyLysGluLeuCysLysMetThrLysAspPheGln 177
Db 65 ACCCTCTTCCGAATATGGATGGCAAGAGCTCTGCAAAACGAACAGGATGATTCCTC 124
QY 178 ArgLeuThrProSerTyrAsnAlaAspIleLeuLeuSerHisLeuHisTyrLeuArgGlu 197
Db 125 CGAACCACTCCCTCTACACACAGAGTTCTGTGTCTCACTCAGTACCTCAGGAA 184
QY 198 ArgGlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIle 217
Db 198 ArgGlyAlaThrPheIlePheProAsnThrSerValTyrProGluAlaThrGlnArgIle 217

```

```

Db 185 AGTAGCTCACTGCTGCTGTAC---AATACTCCATCCACACAGAGCTTCCTCAGCTTT 241
QY 218 ThrThrArgProAspLeuProTyrGluGlnAlaArgArgSerAlaTrpThrSerHisSer 237
Db 242 GCCACCAAGAGGTCTCTCTGTT----- 265
QY 238 HisProThrGlnSerLysAlaThrGlnProSerSerSerThrValProLysThrGluAsp 257
Db 266 -----GCAGGGACACMAAACGTGAACAAGACA-----ACAGAACAG 301
QY 258 GlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSerArgLeuAlaAsn 277
Db 302 CAACGGCTCAGCCAGATCCCTATCAAAATCTGGGGCCCACTAGTAGTCGCTTCCCAAT 361
QY 278 ProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuGluLeuLeuSerAspSerSer 297
Db 362 CTGGGAGTGGGCAGATACAACTATGGCAGTTCTCTCGAGTTGCTGCGACAGTTCC 421
QY 298 AsnSerAsnCysIleThrTrpGluGlyThrAsnGlyGluPheLysMetThrAspProAsp 317
Db 422 AACGCCAGCTGTATCACATGGGAGGGACCAATGGGGAATTCAGATCAGACCCAGAT 481
QY 318 GlnValAlaArgArgTyrGlyGluArgLysSerLysProAsnMetAsnTyrAspLysLeu 337
Db 482 GAAGTGGCAGCGCGCTGGGAGAACGCAAGCAAGCCCACTGAATTAAGCAAGCTG 541
QY 338 SerArgAlaLeuArgTyrTyrTrpAspLysAsnIleMetThrLysValHisGlyLysArg 357
Db 542 AGCGAGCGCTTCGATACTACTATGATGAAGACATTATGACCAAGTGCATGGCAAAAG 601
QY 358 TyrAlaTyrLysPheAspPheHisGlyIleAlaGlnAlaLeuGlnProHisProGlu 377
Db 602 TATGCTACAAATTGACTTTTCATGGCATTGGCCAGGCTCTCCAGCCTCATCCCACTGAA 661
QY 378 SerSerMetTyrLysTyrProSerAspLeuProTyrMetSerSerTyrHisAlaHisPro 397
Db 662 TCATCAATGTACAGTATTCATCAGATCTCTCTACATGCCTTCTTACCATGCCACCAG 721
QY 398 GlnLysMetAsnPheValAlaProHisProProAlaLeuProValThrSerSerSerPhe 417
Db 722 CAGAAGGTGAACTTGTATCCCAACCAACCCCTTCTTATGCGCTGTACATCATCCAGTTT 781
QY 418 PheAlaAlaProAsnProTyrTrpAsnSerProThrGlyGlyIle 432
Db 782 TTTCGAGCAGCCTCACCTTATTGGACCTCCCTGCTGGGAAGCAT 826

RESULT 14
AK083900 2420 bp mRNA linear HTC 20-SEP-2003
LOCUS Mus musculus 12 days embryo spinal ganglion cDNA, RIKEN full-length
DEFINITION enriched library, clone:D130057B15 product:Friend leukemia
integration 1, full insert sequence.
ACCESSION AK083900
VERSION AK083900.1 GI:26350832
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
2
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159

```



```

QY      447 LeuGlyThrTyrTyr 451
Db      837 TTAGGCGAGCTACTAC 851

RESULT 15
BM456833
LOCUS   AGENCOURT 6404041 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5583753
DEFINITION 5', mRNA sequence.
ACCESSION BM456833
VERSION   BM456833.1 GI:18505873
KEYWORDS EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1045)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-k@mail.nih.gov
          Tissue Procurement: ATCC
          cDNA Library Preparation: Life Technologies, Inc.
          DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LAM2347 row: f column: 10
          High quality sequence stop: 569.

FEATURES             Location/Qualifiers
     source           1..1045
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5583753"
                     /tissue_type="embryonal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_92"
                     /note="Organ: testis; Vector: pCMV-SPORE6; Site 1: NotI;
                     Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 2.5 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."

ORIGIN
Alignment Scores:
Pred. No.:      1.82e-79      Length:      1045
Score:          1040.00      Matches:    214
Percent Similarity: 75.24%      Conservative: 17
Best Local Similarity: 69.71%      Mismatches: 33
Query Match:    42.55%      Indels:     44
DB:             12           Gaps:       8

US-09-902-772-2 (1-451) x BM456833 (1-1045)

QY      7 GluAlaLeuSerValSerGluAspGlnSerLeuPheGluCysAlaTyrGlySerPro 26
Db      190 GAAGCCTTATCAGTTGTGAGTGAGGACAGTCGTTGTGTAGTGTGCTACGGAGCGCCA 249

QY      27 HisLeuAlaLysThrGluMetThrAlaSerSerSerGlyTyrGlyGlnThrSerLys 46
Db      250 CACCTGGCTAAGACAGAGATGACCGCGCTCTCTCCAGCGACTATGGACAGACTTCCAAAG 309

QY      47 MetSerProArgValProGlnGlnAspTrpLeuSerGlnProProAlaArgValThrIle 66
Db      310 ATGAGCCACACCGCTCCCTCAGCAGAGATGGCTGTCTCAACCCCGACGACGGGTCAACATC 369

QY      67 LysMetGluCysAsnProAsnGlnValAsnGlySerArgAsnSerProAspCysSer 86
Db      370 AAAATGGATGTAACTTAGCCAGGTGAATGGCTCAGAGAACTCTCTGATGAATGCAGT 429

```

```

QY      87 ValAlaLysGlyGlyLysMetValSerSerSerAspAsnValGlyMetAsnTyrGlySer 106
Db      430 GTGGCCAAAGCGGGAGATGGTGGGAGCCACACACCGTTGGATGAATCACTCGGCAGC 489

QY      107 TyrMetGluGluLysHisIleProProAsnValThrThrAsnGluArgArgValIle 126
Db      490 TACATGGAGGAGAGACACATGCCACCCCAACATGATGACCAAGAGCGGAGAGTTATC 549

QY      127 ValProAlaAspProThrLeuTrpSerThrAspHisValArgGlnTrpLeuGluTrpAla 146
Db      550 GTGCCAGCAGATCTTACGCTATGGAGTACAGACCATGTGGCGGAGTGGCTGGAGTGGCGG 609

QY      147 VallysGluTyrGlyLeuProAspValAspIleLeuLeuPheGlnAsnIleAspGlyLys 166
Db      610 GTGAAGAATAATATGCGCTTCCAGACGTCACATCTTGTATTATCCAGAACATCGATGGGAG 669

QY      167 GluLeuCysLysMetThrLysAspAspPheGlnArgLeuThrProSerTyrAsnAlaAsp 186
Db      670 GACTGTGCAAGATGACCAAGAGAGACTTCCAGAGGCTCACCCCGAGTACACGCCGAC 729

QY      187 IleLeuLeuSerHisLeuHisTyrIleuArgGluArgGlyAlaThrPheIlePheProAsn 206
Db      730 ATCCTTCTCTCACATCTTCACTACCTCAGAGAGACTCTCTTCCACATTTGACTCCAGAT 789

QY      207 ThrSerValTyrProGluAlaThrGlnArgIleThrThrArgPro----- 221
Db      790 -----GATGTTGATAAAGC-CTTACAAACTCTCCCGGTTAATGCATGCT 833

QY      222 -----AspLeuProTyrGluGlnAlaArgArgSerAlaTrpThrSerHisSerHis 238
Db      834 AGAAACACCGATTATACCATATGGACCCCGGAGATCAGCTGGACCGGTACCGTCAA 893

QY      239 ProThrGlnSerLysAlaThrGlnProSerSerSerThrVal-----Pro 253
Db      894 CCC-----CAAGCCCAAGTCGAAAGCTGTTCAAACATTTCTCCCTCAAGGGCCC 944

QY      254 LysThrGluAspGlnArgProGlnLeuAspProTyrGlnIleLeuGlyProThrSerSer 273
Db      945 AAAACTGAA-----AGACCCAGC-----GGCCCTCAGGT--- 974

QY      274 ArgLeuAlaAsnProGlySerGlyGlnIleGlnLeuTrpGlnPheLeuLeuLeuLeu 293
Db      975 -----TAAATCCC-----TTAATCACAATCCTT 998

QY      294 SerAspSerSerSerAsnSerAsn 300
Db      999 TGGGAACCCCAATATAAAC 1019

```

Search completed: May 27, 2004, 23:56:42  
Job time : 3667 secs